Methods and Tools for the Microsimulation of Household Expenditure

Tony Lawson

A Thesis Submitted for the Degree of Ph.D.

Department of Sociology

University of Essex

January 2014

Abstract

Spending by households represents a significant component in the UK economy and the ability to model the effects of socio-economic change on household expenditure is crucial at both the commercial and governmental level. This is usually done by estimating the parameters of a demand system. However, there are several difficulties associated with this approach including, representing the heterogeneity of economic units, the dimensionality of a complex budget set and the specification of the functional form. One way to avoid these is to develop a model that directly simulates the individual units in what is known as a microsimulation. However, models of this type have been found to be complex and expensive to develop. This thesis investigates the possibility of simplifying the development process by using an agent-based modelling toolkit called NetLogo. The idea is tested by constructing a model to project the demographic characteristics of the UK population over time, showing that NetLogo provides a powerful and efficient platform for microsimulation modelling. Then it applies what is known as a random assignment scheme to model household expenditure. This is based on the idea of copying the expenditure pattern from a donor, which is in some sense similar to the receiving unit. Random assignment is then tested by developing a series of models of the effect of demographic and economic change on UK household expenditure patterns. The thesis contributes to methods and tools for modelling household expenditure by developing a framework for the analysis of household spending patterns based on the application of micro-level concepts and techniques throughout. This makes it possible to do what could not be done before which is to have a convenient way to model household expenditure that places no limit on the level of disaggregation or the number of goods represented.

Acknowledgements

This research was sponsored jointly by the UK Economic and Social Research Council (ESRC) and BT. I should like to thank my supervisor Dr. Ben Anderson for his guidance and support. Thanks are due also to Stephen Cassidy and the modelling team at BT, particularly Dr. Morgan Potter for providing the data used in the energy price change model. I thank my wife, family and friends for their support during this time.

Table of Contents

List of Figures

Index of Tables

Chapter 1: Introduction

1.1 Overview

For governments, the ability to anticipate how consumer spending is affected by economic conditions, demographic change or its own policy initiatives, is a key component in managing the economy. Also, for commercial organisations, projecting changes in the level of demand for their goods and services, in a changing world, can make the difference between success and failure.

This is usually done using econometric methods that involve estimating the parameters of an equation that links spending patterns with a set of variables that are thought to influence spending behaviour (Clements and Hendry, 2002). However, as this thesis argues, the econometric approach is limited in its ability to allow for the way that the differences between individual cases affect the way spending patterns change in response to varying socio-economic conditions. One solution is to develop a model that directly simulates the individual units in what is known as a microsimulation, as proposed by Orcutt (1957). The aim of this research is to develop a conceptual and technical framework for the microsimulation of household expenditure and then show how it can be used to model the effect of socio-economic factors on household spending patterns.

Based on a comprehensive review of current methods, a number of problems or gaps in the available technology are identified. The research then brings together and tests

new combinations of methods and tools by constructing and evaluating microsimulation models of how household expenditure patterns are affected by income and demographic change.

This introduction begins with an overview: first of microsimulation modelling and then expenditure analysis. In each case it identifies problems or difficulties with current methods and these provide the motivation and justification for this research. This leads on to a definition of the research questions that the thesis is intended to answer. The chapter concludes by outlining the structure of the main body of the thesis.

1.2 Microsimulation Modelling

Microsimulation (Orcutt, 1957) was conceived to solve a general problem in macro-economic modelling, which was that the aggregated variables used to represent the parameters of an economic system do not correctly represent the distribution of those parameters as measured at the level of individual decision making units. Orcutt illustrated this with a numerical example where there is a set of individuals which are imagined to have input X and output Y as shown in the table below.

X	**Y**
0	0
1	1
2	1

It can be seen that a population containing 100 individuals, all of which have an input

of 1, will generate an output of 100. It is also apparent that if the population is made up of 50 individuals with an input of 0, and another 50 with an input of 2, the combined input will still be 100 as before but the sum of the output will now be 50. Thus, knowledge of inputs at the macro-level does not uniquely determine the output of the system as a whole. Orcutt proposed to solve this problem with a model that preserves the micro-level characteristics of the individual elements. This takes the form of a simulation composed of 'interacting units' which receive inputs and generate outputs. The inputs he defines as anything which acts upon, or is taken account of, by the unit. These could be things like age, income or the rate of inflation. They could also be the outputs from other microsimulation units. Outputs are anything which stems from, or is generated by the unit. Examples of outputs include marriage or the birth of a child. The heart of the unit that generates outputs from inputs is what Orcutt calls the 'operating characteristics'. These can be implemented in a variety of ways including equations, graphs or tables.

The first microsimulation model was implemented as a computer program in 1961 when Orcutt, with the help of some of his students implemented DYNASIM to model the socio-economic status of individuals and families (Orcutt et al., 1961). In the 1970s, he developed a more sophisticated version called DYNASIM2 (Orcutt et al., 1976), that was used for several years to model the financial situation of retired people in the United States. Dynasim3 represents an updated version using new data sources. The DYNASIM structure, which includes modules covering births, deaths, partnership formation and dissolution, has served as a template for many subsequent

implementations (Spielauer, 2007). These include MOSART in Norway, developed to model public expenditure (Fredriksen, 1998) and DESTINIE in France, to project pension requirements (INSEE, 1999). In the UK, a range of models has been developed to assist in policy formation. These include PENSIM, which has been used in pension forecasting (Hancock et al., 1992), CARESIM for anticipating demand for healthcare (Zaidi and Rake, 2001) and SAGE (Simulating Social Policy in an Ageing Society) (Cheesbrough and Scott, 2003) which simulates the effects of demographic change on demand for pensions and healthcare.

Currently, there are several varieties of microsimulation in use (O'Donoghue, 2001a), (Li and O'Donoghue, 2013) but in general they can be divided into two main types: static and dynamic (Merz, 1993). In a static microsimulation, the unit generates a set of outputs from a set of inputs but is not itself changed in the process. In a dynamic microsimulation the unit changes or 'ages' over time. This makes dynamic microsimulation useful for modelling long-term processes such as demographic change. This thesis is primarily concerned with dynamic microsimulation.

By the late 1990s, microsimulation was making a contribution to informing government policy in several countries but the models were found to be complex and expensive to develop. A major factor in the high cost of microsimulation modelling has been in developing the software needed to run them (Harding, 2007). This research will investigate one way to alleviate this problem by using an agent-based modelling (ABM) toolkit. These seem to have much of the functionality of a

microsimulation model already built in, such as mechanisms to handle the collection of agents and an advanced graphical user interface (GUI). As such, they might well have the potential to reduce the complexity of developing the model. While there are already some examples of microsimulation models developed using an ABM platform such as LaborSim (Leombruni and Richiardi, 2005) and IFSIM (Baroni, Zamac and Oberg, 2009), these have all been of a type that utilises a general-purpose programming language and this still requires advanced programming skills on the part of the developer. There has been no attempt to apply a simpler, script-based ABM platform. One of the objectives of this research will be to investigate the extent to which one of these toolkits, known as NetLogo, can reduce the burden of developing a dynamic microsimulation model. This will be done by implementing one such model using NetLogo and evaluating its performance in terms of its ability to implement all the necessary functions, its processing speed and usability. The resulting dynamic microsimulation model, known as Tyche, will then be combined with an expenditure component and applied to model household spending patterns.

1.3 Expenditure Modelling

In 2009, consumer spending in the UK amounted to 872 billion pounds (ONS, 2010a). Figure 1 shows how this spending was distributed among the 12 high-level categories of the annual Expenditure and Food Survey (EFS).

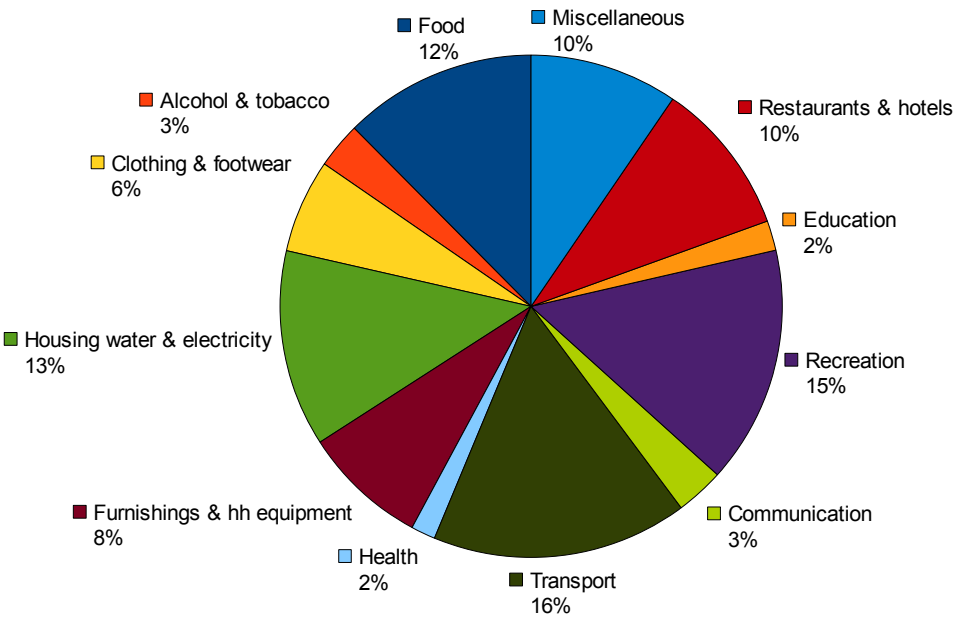


Figure 1: Distribution of UK Household Spending

These figures, taken from the 2006 EFS, relate to consumption expenditure which can be thought of as spending on things to *use* rather than sell or trade. Commercial organisations operate in one or more of these sectors and the share of household expenditure devoted to each one gives an indication of the size of the market for a particular good or service. Market size is known for the current and previous years but it is desirable to have some understanding of the factors that influence spending and then to project it into the future. This might assist, for example, in an organisation's strategic planning. If a market is projected to grow then it might be worth expanding activities in that area or investing in new technology. If the sector seems likely to decline then it may be advisable to consolidate or perhaps look to another sector.

Governments obtain revenue from taxation and this affects how much households have remaining to spend on consumption items. This is particularly relevant in the case of indirect taxes where the level of taxation on a particular item can affect how much of it is consumed and as a consequence, affect the amount of revenue generated.

The importance of analysing spending patterns for such a wide range of organisations has led to the development and application of a formidable array of methods. These range from approaches that rely on the knowledge and judgement of individuals to complex mathematical models. Some are essentially macro methods in that they model aggregated variables and the relationship between them. When modelling expenditure patterns at the level of individual households, the dominant approach is to use econometric methods (Clements and Hendry, 2002). It is possible to do this in a single, regression type equation of the form:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$$

Here, the dependent variable Y might represent the budget share for food. The independent variables X_1 to X_n could represent factors that are thought to influence spending on food such as household size, income and price. The constants b_0 to b_n would be estimated using standard statistical software on observed data that captures the relationship between the relevant variables.

One of the problems with this approach is that it is necessary to specify a separate equation for each good of interest. This becomes unwieldy if the number of goods is

large as it would be in a typical household budget set. It is also difficult to model the interaction between spending on each good because, in principle, this will depend on what is spent on all the other goods. The list of independent variables should then include the budget shares of all these items. This is feasible for a small number of goods but as the size of the budget set increases, the number of parameters needed to estimate the model grows quickly to the point where, for most datasets, there would not be enough cases to provide accurate estimates of the parameters.

This problem is alleviated to some extent by the use of complete demand systems consisting of an integrated set of equations. Some of the most sophisticated are the Quadratic Almost Ideal Demand System (QAIDS: Banks, Blundell and Lewbel, 1997) and the Exact Affine Stone Index (EASI: Lewbel and Pendakur, 2009). They use the principles of neoclassical economic theory to impose restrictions on the possible values of the parameters and so reduce the amount of data needed to estimate them. However, this does not solve the problem completely because the number of parameters to be estimated still rises with the size of the budget set and this places a limit on the number of goods that can be modelled at any one time. Also, more significantly, in the context of modelling household expenditure, it will be demonstrated later that the number of parameters to estimate increases with the number of *combinations* of household attributes. Hence, in order to reduce the number of parameters, it is necessary to aggregate households with similar characteristics into homogeneous types. As a result, even if the base data represents the population at the level of individual units, the aggregation of goods and

households into groups means that the model represents a set of averages for the items modelled and what is estimated in a demand system is the average demand conditional upon observed characteristics (Christensen, 2007). In that case, change within the groups is not represented and any output can only be done in terms of the previously defined classifications. This runs counter to the central principle of microsimulation which is to work with the individual cases and makes it impossible to represent the full heterogeneity of households as they change over time.

There is an alternative method, known as random assignment, that was described by Klevmarken (1997) and evaluated in the context of modelling geographic mobility by Holm, Mäkilä, and Lundevaller (2009). This is based on the idea of obtaining unknown variables, analogous to the dependent variables of an equation, by copying from a donor case that has similar characteristics. According to Klevmarken (1997), the advantages of this method are that it is not necessary to impose a functional form on the data or make any assumptions about the distribution of variables. There are no parameters to estimate and the method preserves the variation and most of the correlation present in the original dataset. The idea of copying from a similar case seems to present a method of modelling household expenditure that is compatible with the defining characteristic of microsimulation. Also, the absence of parameters and any need to make assumptions about the distribution of cases, appear to make it possible to retain all the useful micro-level information that is derived from the input data set.

In this thesis, random assignment is adapted to model household expenditure patterns so that when a household changes its characteristics, another household which is similar to the one with the new configuration is selected and its expenditure pattern is copied onto the subject household. The research evaluates random assignment in this new application area in terms of feasibility (whether and how conveniently it can be done), validity (whether it produces the expected results) and flexibility (the range of situations to which it can be applied).

1.4 Research Questions

The preceding two sections outlined the general area of the research described in this thesis. Then, they identified some areas that can be seen as problematic for users of current methods and went on to suggest alternative approaches that, it was claimed, would ameliorate or avoid these problems. The aim of this thesis is to test these claims and can be stated in the form of the question:

> *How can random assignment and NetLogo be combined to develop a coherent micro-level framework for the analysis of household expenditure patterns?*

The justification for this research is partly theoretical in that current methods of expenditure analysis rely on demand systems which were developed within an economics tradition, independent of microsimulation. Klevmarken (1997) notes that this limits the representation of behaviour to what is the current state of the art in economics. It can also be said that it limits the analysis to the conceptual framework,

methods and techniques that are characteristic of the economics paradigm. This research uses micro-analytical concepts throughout to develop a coherent micro-level framework for expenditure analysis. There is also a strong practical component to the justification in that, since random assignment operates at the micro-level, it will be possible to retain the distribution of cases as the system changes. Also, since the copying process can access all the variables in a donor case, there will be no limitation on the number of goods that can be modelled simultaneously.

The presence of NetLogo and random assignment in the overall research problem implies that it has two components, one is to do with applying ABM techniques to microsimulation, the other is to test the suitability of random assignment in the context of modelling household expenditure. As a result the main research question can be resolved into two parts:

> *1. How suitable is NetLogo as a platform for developing a dynamic microsimulation model?*
>
> *2. To what extent can random assignment form the basis of an approach for expenditure analysis at the micro-level?*

The significance of the first question is that if a script-based ABM platform like NetLogo is found to reduce the burden of programming such models, dynamic microsimulation may become accessible to a wider group of users than has previously

been the case. The second question is significant because if random assignment is shown to be a practical and valid approach for expenditure analysis, it will be possible to do what cannot be done currently, which is to model an unlimited number of goods simultaneously at a disaggregated level where the output can be produced at any desired level of detail with no loss of information due to aggregation.

1.5 Methods

These questions will be approached in a practical manner by developing a set of microsimulation models using NetLogo. This begins with a dynamic microsimulation model to project demographic change in the UK population. The model, known as Tyche, is validated using methods described by (Morrison, 2008). The process of developing the model facilitates answering the question of NetLogo's suitability as a platform for microsimulation. This is augmented by comparing Tyche against another microsimulation framework called LIAM2, in terms of functionality, processing speed and usability.

The second model to be developed using NetLogo is to test the random assignment scheme as a method of expenditure analysis. This is done by creating a model to project changes in household expenditure patterns in response to changing income. The results are checked against stylised facts derived from previous literature on the relationship between household income and expenditure patterns. The model is then used to demonstrate how random assignment can be applied to generate disaggregated results showing how different sections of the population respond to

changing income levels. This process is directed towards addressing the questions of the feasibility and validity of random assignment as a way of modelling household expenditure. The flexibility of random assignment is tested in a further substantive application which combines the dynamic microsimulation with a random assignment scheme to model the effect of demographic change on expenditure patterns.

The impact of the ageing population on spending patterns has usually been studied using a range of econometric methods, which, as argued above, makes it difficult to model the distribution of cases as they change over time. This work has also been hampered by a reliance on official demographic projections, which were not designed to provide the most appropriate parameters for expenditure analysis. The combined demographic and expenditure model is used to obtain the relevant parameters to project spending patterns in the UK to the year 2036. The results show that, allowing for UK Office for National Statistics (ONS) assumptions of future changes in birth and death rates, demographic change, *ceteris paribus*, would to lead to an increasing population and so to increased demand in most expenditure categories. As such, in addition to demonstrating how NetLogo and random assignment can be used to model household expenditure patterns, these results contribute to the debate on the socio-economic effects of population ageing.

Following this, two further applications of the random assignment technique are described, which were developed in collaboration with the co-sponsor of this research, BT. The first one models how consumer's spending patterns respond to

changes in the price of energy, the other investigates the effect on household budgets if significant numbers of people were able to work from home. In particular, where would the money saved from not commuting be spent?

This method of practical investigation allows the two general questions formulated above to be decomposed into a number of more specific questions which can be approached objectively:

1. Does NetLogo have sufficient functionality to implement all the usual functions of a dynamic microsimulation model?

2. Does NetLogo have any additional features that are not usually found in existing dynamic microsimulation models?

3. Is its processing speed adequate for the task and how does it compare with an established example?

4. How easy is NetLogo to use: a) from a developer's point of view in creating a new model and b) from an end user's point of view, running a model and obtaining results?

5. Is it feasible to use a random assignment scheme for modelling household spending patterns and are there any difficulties to be overcome?

6. Can a model, implemented using a random assignment scheme, produce standard results that would be expected from previous research?

7. Is random assignment applicable in a wide range of areas and does it have any limitations?

The answers to these specific questions will inform the response to the two more general questions. The response to these will provide evidence for or against the main assertion that NetLogo and random assignment can from an effective approach for modelling household expenditure.

1.6 Structure of the Thesis

This introduction gave an overview of micro-level modelling methods and expenditure analysis to support the thesis that current methods of modelling household expenditure have some limitations that can be alleviated by the use of NetLogo and random assignment. The next two chapters broaden and deepen this argument by providing a review of literature on micro-level modelling and expenditure analysis. The literature review phase is quite extensive because, unlike many Ph.D. theses, which focus narrowly on one topic, this thesis draws on three distinct areas: microsimulation, agent-based modelling and expenditure analysis.

Chapter 2 begins with a more detailed description of microsimulation technology, including how it works, why it is used, what it is used for and how it is implemented.

Some of the difficulties associated with microsimulation modelling are discussed next. These are a set of issues that are currently problematic and represent important areas for research in the 21st century including: the complexity of developing microsimulation models, their high cost and development time, poor usability, difficulty of validation, lack of behavioural representation, poor accessibility for new modellers and a lack of predictive power. The particular problem that this thesis focuses on is the complexity involved in developing a dynamic microsimulation model and it is argued that addressing this issue, by using an ABM toolkit, makes an indirect contribution to reducing problems in several of the other areas. The various types of ABM toolkit are reviewed next and it is suggested that NetLogo offers the greatest potential for reducing the burden of software development. Chapter 3 reviews the literature on expenditure analysis. It describes the development of progressively more sophisticated approaches culminating in the QAIDS and EASI models and their variants. Then it highlights some of the difficulties in modelling change at the micro-level using current methods and proposes random assignment as an alternative.

Following this, the next two chapters investigate the solutions to the problems identified in the literature review, in the form of NetLogo, to make microsimulation easier and random assignment for modelling expenditure at the micro-level. Chapter 4 describes the construction and validation of a dynamic microsimulation model to test how suitable NetLogo is for this task. Chapter 5 uses a random assignment scheme to model the effect of changes in household income on expenditure patterns, to test the feasibility and validity of random assignment in this application.

Chapter 6 combines the NetLogo demographic model with expenditure modelling using a random assignment scheme, to model the effects of structural population ageing on household expenditure patterns. This demonstrates the application of what is intended to be a coherent micro-level framework for the analysis of household expenditure patterns. It shows how the method can be used in a substantive application and contributes to the debate on this important topic. It finds that population ageing, isolated from other influences, would lead to an increase in demand for most expenditure categories and suggests that the additional economic activity has the potential to be beneficial to the wider UK economy. Chapter 7 demonstrates the use of the random assignment method in a commercial environment, describing the 'energy price change model' and the 'homeworking model'.

Chapter 8 concludes by reviewing the research described in this thesis. It begins by using the information gained in developing the models, to answer the research questions posed above; beginning with the detailed questions and working towards the main question. Next, the significance of the findings is discussed in terms of their contribution to knowledge in the field of microsimulation. This is based on noting where they have been disseminated, such as in peer reviewed journals, book chapters and working papers, also the extent to which they have been made use of by others, including the co-sponsor, BT. Following this, the significance of the research is discussed in the light of the problems that were identified in Chapter 2. The potential for further work is indicated next and the thesis concludes by highlighting the way in which the combination of NetLogo and random assignment increase the scope of

methods and tools for the microsimulation of household expenditure.

1.7 Conclusion

This chapter has described the nature of the research to be presented in this thesis; the reasons for doing it and why it claims to be novel and significant. The ability to model the effects of social, demographic and economic change on the demand for goods and services is crucial at both the commercial and governmental level. While a range of tools have been applied in this area, the motivation for the research lies in the perception that these are lacking in some regards. In economic modelling the difficulty is in representing the way the heterogeneity of individual units affects the marginal variation in spending patterns as conditions change. In microsimulation, the problem has been in the complexity of developing the software. Essentially, the research presented here involves applying existing technology, in the form of NetLogo and random assignment, in a new area and evaluating the extent to which the perceived problems are solved. This is begun by implementing and testing a dynamic microsimulation model using NetLogo and a model of the effect of changes in household income on expenditure patterns. Then the two components are brought together in a substantive model of the effect of population ageing on UK aggregate spending. This informs the debate on what has been a central topic for microsimulation modelling by showing that population ageing is likely to increase demand for most expenditure categories and suggests that this has the potential to bring some economic benefit. However, as the further applications developed for use at BT demonstrate, the research presented here goes beyond the study of one

application of a microsimulation model. Rather, it aims to develop a framework comprising NetLogo and random assignment that can be used to tackle class of problem where it is useful to retain information on the distribution of cases. As such it hopes to lay the foundation of a method that can be utilised and built upon in the future.

Chapter 2: Micro-level Modelling

2.1 Introduction

This chapter reviews current methods and tools for modelling at the micro-level. It begins with a description of microsimulation, covering the range of different types, how they work and what they are used for. This leads on to the identification of one of the problems that this thesis addresses, which is the complexity of implementing a dynamic microsimulation model. It is proposed that the inbuilt features of agent-based modelling (ABM) toolkits offer a way to reduce the burden of developing the software. Previous reviews of ABM toolkits are used to guide the selection of which one is likely to provide the most assistance in this respect. The reasons for choosing NetLogo are given and its main features are described. The chapter concludes by examining the range of current problems for microsimulation modelling. These provide a baseline, which will be used in the concluding chapter, as a way to evaluate the significance of the contribution this thesis makes to microsimulation modelling.

2.2 Modelling and Simulation

The precise nature of modelling and simulation are a matter of some debate. Hartmann (1996) for example, defined modelling, in the widest sense, as a set of assumptions about some system, and simulation as an imitation of one process by another process. For Maria (1997) a model is a representation of some system of interest which should be a close approximation of the most salient features of the system but not be too complex to understand and experiment with. Simulation is then

defined as the operation of the model which may be used to predict changes to the system or it may be used to perform experiments that are not practical with the real system. For Gilbert and Troitzsch (2005) a model is a simplification which is smaller, simpler and less complex than the target system it represents, while simulation is a particular kind of modelling which can be used for explanation and prediction. When referred to in this thesis, a model will be taken to be an abstraction of selected features of a system. A simulation will be an experiment that involves running the model on a set of data with the aim of predicting or explaining the effect of changes in one set of features on another set of features.

Models have been constructed using a range of media. These have included mechanical models, such as an orary, mathematical models consisting of a set of equations and even a fluid based model of the UK economy (Hayes, 2009). While this was a kind of analogue computer, it is only quite recently that digital computing has been capable of developing useful simulation models. Gilbert and Troitzsch (2005) provide an account of the range of methods that are currently applied in social simulation. These include system dynamics (Forrester, 1961), queueing models (Kheir, 1988), microsimulation (Orcutt, 1957), cellular automata (Wolfram, 2002) and agent-based modelling (Railsback and Grimm, 2011). Edmonds and Meyer (2013) provide a comprehensive account of modelling social systems.

The various modelling approaches can be divided into two main types: macro and micro (Gilbert and Troitzsch, 2005). Macro models use aggregated variables to

represent some property of the system and often make use of mathematical equations to describe how the variables interact. This approach can be used in situations where there is comparatively little data on the system which means it usually has light computational requirements. It also allows a formal, unambiguous description of the system. However it is difficult to model non-linear dynamics, partly because of the difficulty of working with non-linear functions and also because a small element in the system, below the level of resolution of the model, can sometimes have a disproportionate influence on the evolution of the system as a whole.

Micro-level approaches are characterised by the representation of the individual components of the system and the way they interact. This places high demands on data availability and computing power because every element of the system must be described and processed. They have the advantages of being well suited to modelling non-linear behaviour, can produce disaggregated output (for each element if necessary) and they can also represent the effect of the distribution of characteristics. As discussed in Chapter 1, it was primarily this last advantage that prompted Orcutt to propose microsimulation modelling.

This literature review now focuses more closely on two of the micro-level methods mentioned above which have particular relevance to the thesis. Firstly microsimulation and then agent-based modelling.

2.3 Microsimulation

A microsimulation is a model which uses simulation techniques and which takes micro-level units as the basic units of analysis when investigating the effects of social and economic policies (O'Donoghue, 2001). Since Orcutt's early insights, mentioned in Chapter 1, the advantages of modelling at the level of individual units have been recognised and applied in several areas. Some of the main types of model are discussed briefly here.

2.3.1 Traffic Microsimulation

In contrast to the traditional method of modelling traffic in terms of aggregate flows, traffic microsimulation represents the behaviour of individual vehicles (Gipps, 1981). This makes it possible to model the interaction of vehicles at complex road junctions and reproduce emergent phenomena such as the appearance of shockwaves in dense traffic. Examples of traffic microsimulation models developed in the UK include DRACULA (University of Leeds), PADSIM (Nottingham Trent University), PARAMICS (the Edinburgh Parallel Computing Centre and SIAS Ltd.), SIGSIM (University of Newcastle) and SISTM (Transport Research Laboratory). Algers et al., (1997) provide a comprehensive review of these and many other traffic microsimulation models worldwide. Ravulaparthy and Goulias (2011) provide a wide ranging review of traffic, land use and demographic microsimulation modelling methods and applications.

2.3.2 Spatial Microsimulation

The base data for microsimulation modelling is often derived from a large-scale social survey but for very good confidentiality reasons, these do not provide the precise location of each household. Spatial microsimulation modellers have developed a range of techniques to merge aggregated sub-regional census data with the detailed household characteristics provided in survey data (Birkin and Clarke, 1988; Williamson et al., 1998; Voas and Williamson, 2000; Ballas et al., 2005a; Chin et al., 2005, 2007; Lymer et al., 2008). The knowledge of population characteristics at the local level is used in research into the effects of the spatial variation of demographic and economic characteristics. The synthetic datasets generated in spatial microsimulations can be used directly such as where Jing et al. (2014) developed a travel choice model to estimate CO2 emission at the micro-spatial level in Beijing. However, in many cases, the base data set is updated over time to represent change in local areas. SimBritain (Ballas et al., 2005b) models UK population demographics up to the year 2021 by combining data from the British Household Panel Survey (BHPS) with census data. Anderson, De Agostini and Lawson (2014) combined a dynamic microsimulation model, a demand system and spatial microsimulation to investigate the effects of economic austerity measures at the small area level. Another important application area for spatial microsimulation is in modelling travel demand (Goran, 2001). Tanton and Edwards, (2013) and Heppenstall et al., (2013) provide comprehensive guides to spatial microsimulation.

2.3.3 Static Microsimulation

This type of microsimulation has been applied in many areas but it is most prominent in modelling changes to the tax and benefit systems in many developed countries.

2.3.3.1 Arithmetical Microsimulation

Sometimes known as micro-accounting models, arithmetical microsimulation is, in a sense, conceptually less complex when compared to other types of microsimulation model. This is due largely to the assumption that the micro-unit's behaviour is constant and therefore does not need to be modelled explicitly. Nevertheless, representing the rules of a tax-benefit system is certainly a highly complex task. Despite this, most developed countries have implemented at least one such model to gain a detailed understanding of how their tax-benefit system affects all sections of the population, as well as to understand the effects of any proposed change in regulations. One model in the UK of this type is POLIMOD (Mitton and Sutherland, 1999). Most of the input data for POLIMOD is derived from the UK Family Expenditure Survey. This is supplemented with data on self employment, income tax, mortgage rates and population weights. A SAS (Statistical Analysis System) program is used to convert these data into a set of ASCII files at the individual, family and household levels, as well as a household level expenditure file. A set of parameter files is constructed containing all the rates for tax, and benefits both before and after the proposed policy change. These include rates for child benefit, national insurance, income tax, council tax, VAT, interest rates and excise duty. In addition, data on retail prices are used in the calculation of each household's excise duty liability. The input

files and parameter files are then processed by the POLIMOD program, which was written in C, to calculate each unit's income and expenditure before and after the policy intervention. The distributional output is then produced by another C program (Mitton and Sutherland, 1999).

A further step up in complexity has been undertaken in the development of EUROMOD (Sutherland and Figari, 2013) which is intended to represent the tax-benefit systems of all EU countries. However, it can equally well be adapted for non EU nations as was done for South Africa (Wilkinson, 2009) and a set of Latin American countries (Absalon et al., 2009). To accomplish this level of versatility, it was designed in a modular form where the function of each part of the tax-benefit system is separated from national parameters. In this way, no part of each national system is hard coded into the program and it is possible to transfer parts of systems between nations. This allows comparative analysis and the effects of 'policy swapping' between nations to be investigated. EUROMOD is the most widely used static tax-benefit model in the world and the only UK model that is generally available from the developers (Williamson, 2009).

2.3.3.2 Behavioural Microsimulation

Arithmetical microsimulation models stop when they have calculated how much better or worse off a micro-unit is after a policy change. Behavioural microsimulation models attempt to predict what the unit does in response to being better or worse off. This is done on the assumption that the unit will act to maximise its utility under the

new budget constraint. A model can then be specified where an equation is defined, linking all the variables of interest (Bourguignon and Spadaro, 2006). Next, the parameters are estimated and the resulting equation allows the behavioural response to be determined by varying the independent variables in the equation. A common application is to predict changes in labour supply following changes to the tax-benefit system. One example of this is the Institute for Fiscal Studies' labour supply model, Simulation Package for the Analysis of Incentives (SPAIN: Duncan, 1991) which uses results from its tax and benefit model (TAXBEN: Giles and McCrae, 1995) to examine behavioural responses to policy changes.

2.3.4 Dynamic Microsimulation

The policy change that is to be modelled in a static microsimulation might be planned to take effect at some time in the future; changes to the pension system for example. During the intervening time, it is possible that the nature of the population might change in some way. The rate of unemployment, age distribution or household income might all vary. In this case, it would be desirable to alter the current population so that it corresponds to the way it is expected to be in the future. The process of updating the population over time is called ageing.

2.3.4.1 Static Ageing

In static ageing, there is no attempt to represent the internal processes that drive demographic change. Instead, population or group aggregates are aligned to conform to values from an external source. Shifts in the nature of the population will change

the probability of observing a micro unit in a particular state. This can be simulated by re-weighting the original sample according to external forecasts. As the original weights contain a good deal of information about the sample, it is desirable to minimise the overall change in their values. This can be done using one of a variety of distance functions. Software is available to assist in re-weighting such as Clan97, developed by Statistics Sweden (Andersson and Nordberg, 1998).

2.3.4.2 Dynamic Ageing

In dynamic ageing, the objective is to simulate as accurately as possible, the processes that drive demographic change. In dynamic microsimulation modelling, demographic projections are done by applying a transition probability to each unit. As the individual microsimulation elements change state, the make up of the population shifts over time. However, this approach to population modelling is not the only way it can be done and outside the microsimulation community, as Imhoff and Post (1998: 97) note, population projections are 'almost invariably' produced using what is known as the cohort-component method. This operates by multiplying a matrix containing the numbers in each age group by another matrix, called the Leslie matrix (Leslie, 1945, 1948), which contains the period specific fertility and mortality rates. This process continues for as many time periods as required. In this method, the demographic system is interpreted as a discrete space Markov process where the probability of being in a particular state depends entirely on the previous state (Andreassen, 1993). In this case, the Leslie matrix can be interpreted as a table of transition probabilities and this opens up a way of performing the calculation in a

different way. Instead of applying mean transition rates to groups, the transition probability can be applied to each member of the population. This is implemented by drawing a random number from a uniform distribution in the range 0 to 1 and comparing its value to a threshold probability from the matrix. If the random number is less than or equal to the threshold, the event occurs, otherwise the unit is unchanged. Since the result of the draw for each member of the population is independent of the result of the draw for other members, due to the central limit theorem (Adams, 2009) the sum of the transitions will converge to a normal distribution (Andreassen, 1993). It is therefore straightforward to obtain the variance and confidence interval for the average number of transitions such as births, deaths etc. This approach is used later to estimate the amount of random variation between each simulation and so measure the component of uncertainty in the model that is due to statistical variation.

At first it might appear that the introduction of stochasticity is an unnecessary complication because it adds uncertainty to the results but Andreassen (ibid) finds six advantages for what is now a microsimulation method, over macro-level modelling.

1) In a microsimulation, the number of elements of data it is necessary to keep track of is equal to the number of individuals in the population multiplied by the number of attributes each one has. For example, if there are 1000 individuals in the population and each has five attributes then there will be 5000 items of data. In the matrix multiplication method, the number of elements is the product of the number of

attributes and the number of classes in each attribute. In this case, there might be an attribute 'age' divided into 20 five-year groups, another attribute, 'sex' with two classes, and 'housing tenure' with three types so there will be 20 * 2 * 3 = 120 elements. This approach seem promising until more attributes are needed. 'Income' might require 10 levels, 'employment status' 5 and 'car use' another two. There will now be 20 * 2 * 3 * 10 * 5 * 2 = 12,000 elements. As further attributes are added, the list is multiplied with every new addition so that the matrix multiplication method quickly becomes impractical. This gives rise to an important advantage of microsimulation in that the model can be expanded to include a large number of attributes.

2) Individual life histories of members of the population are modelled. It is thus possible to map individual trajectories over time and distributional output can be very flexible.

3) The interaction between individuals can be modelled, such as partnership formation and household composition. This means it is easy to ensure the output is consistent, such as there always being the same number of males who marry as females.

4) The ability to follow each individual through their life cycle makes it relatively straightforward to change behavioural assumptions and run alternative scenarios.

5) Microsimulation models make it possible to allow for unobserved heterogeneity within the population under consideration. To do this, it is first necessary to derive a probability distribution between observed variables that reflects the effect of the unobserved variables. When this is done, individuals can be assigned characteristics drawn from the probability distribution.

6) Lastly, microsimulation is an intuitive approach that is easy to understand.

Imhoff and Post (1998) also find a number of advantages for microsimulation over the more usual cohort-component method. Prominent among these is that microsimulation can provide a much richer output. Since microsimulation preserves the individual units, the output can be presented in any form desired, without re-specifying the component groups. Microsimulation is better suited to modelling any interaction between variables and units. It is also capable of modelling continuous variables without dividing them into bands.

Microsimulation is not without its disadvantages, such as high development costs, data requirements and the difficulty of comprehensively validating a complex model. However, for the purposes of this thesis, there is one decisive factor which precludes the use of the cohort-component method in favour of a transition probability based approach. The essence of microsimulation modelling is to work with individual cases in order to *preserve information on the distribution of variables*. The cohort-component method entails defining groups or cohorts, which is incompatible with

microsimulation and this is the reason why transition probabilities are used in this thesis, instead of the standard cohort-component method.

2.3.4.3 Alignment

One feature of dynamic microsimulation modelling that has been particularly controversial is the issue of alignment. It is often the case that long-term demographic projections differ unacceptably from official projections or what has been observed in past data. In this situation, alignment can be used to ensure that model outputs more closely approximate key benchmarks. This is done by adjusting the likelihood of events, for each individual unit, in such a way that the aggregate number of events for some key parameter or parameters is acceptably close to observed or target values. Li and O'Donoghue (2014) describe several methods of implementing alignment such as multiplicative scaling (Neufeld, 2000), the central limit theorem approach (Morrison, 2006) and sort by the difference between logistic adjusted predicted probability and random number (SBDL: Flood et al., 2005). The main justifications for alignment are to make model outputs correspond to target values and so gain credibility among policy makers as well as to ensure compatibility with the macro results from other agencies (Harding, 2007). Anderson (2001) notes that almost all existing dynamic microsimulation models are adjusted to align to external projections of aggregate or group variables when used for policy analysis.

One of the hazards of using alignment is that the outputs will still match target values even if there is a structural change in the process generating the data (Li and

O'Donoghue, 2014). Also, there is a question of consistency and the level of disaggregation at which alignment should take place (Baekgaard, 2002). It also limits the usefulness of the microsimulation as a predictive tool because the model projections are overridden by the alignment process (Harding, 2007). One of the most vociferous critics of alignment has been Winder (2000) who finds the practice to be an 'indefensible fiddle' which means the model cannot be tested properly and suppresses the possibility of emergent phenomena arising in the model.

One of the reasons that microsimulation models are often poor predictors of future macro parameters is that they are necessarily, estimated using past data. This is limited by sample size and duration over which it was collected so it can only model relationships that fell within its observation window. These will be affected by period and cohort effects. As the likelihood of events change in the real world, there is no mechanism (apart from alignment) to update the model specification to allow for this. Also, the model is a simplification of the real system and does not capture all possible demographic types or modes of household formation or feedback effects as individuals respond to macro-level socio-economic conditions.

Winder's alternative to alignment is to work towards a state where microsimulation models give accurate results without alignment and to make explicit links between macro and micro dynamics. Whether this is likely to provide a general solution to the credibility and usefulness of microsimulation modelling can be gauged by noting how microsimulation models are applied in practice. The principle advantage of

microsimulation is to represent the distribution and heterogeneity of cases at the micro-level as they evolve oner time. As we have seen in Chapter 1, macro-level models, cannot do this as effectively. This is why Orcutt (1957) developed microsimulation. Conversely, there are several methods of forecasting which, in many cases provide better predictions than microsimulation. Time-series analysis (discussed in the next chapter) for example, provides an array of powerful techniques for extrapolating from data collected over a period of time. This makes it more suitable than microsimulation for capturing long term dynamic change but it cannot represent the distribution of cases. In this context, it seems that alignment provides a way to combine the forecasting capabilities of other methods with the ability to model at the micro-level so that an aligned microsimulation model is both a good predictor and represents micro-level distributions. The use of alignment also extends the applicability of a dynamic microsimulation model from unconditional forecasting of some future state to generating a series of conditional projections based on a range of counterfactual scenarios. As such, the microsimulation model becomes an experimental platform for explanation as well as prediction, by modelling the causes of observed distributions under a range of different assumptions. However, further research into alignment techniques is required, for as Li and O'Donoghue (2014) note, there is a lack of studies analysing how projections and distributions change as a result of the use of different alignment methods.

2.4 Applications of Dynamic Microsimulation Modelling

The application areas of traffic, spatial and static microsimulation have been

discussed briefly above. Since this thesis develops and uses a dynamic microsimulation model, the applications of this approach are considered in more detail.

Orcutt originally proposed dynamic microsimulation as a way to 'investigate what would happen given specified external conditions and governmental actions' (Orcutt, 1957, 122). His DYNASIM III model was subsequently applied in that vein to evaluate the effect of various tax, social insurance and pension policies. Since that time, the range of applications has widened to include: kinship relationships (INAHSIM: Inagaki, 2010), economic growth and inflation (MOSES: Eliasson, 1977), land use (SustainCity: Morand et al., 2010) and the effect of human systems on the environment (SVERIGE: Vencatasawmy et al., 1999). Despite this, as a comprehensive survey of dynamic microsimulation models shows (Li and O'Donoghue, 2013), the core of microsimulation is still concerned with policy modelling, particularly in the areas of tax, benefits, and pensions; although there has been an expansion into the areas of: education (GAMEO: Courtioux et al., 2008), labour market participation (Kalb and Thoresen, 2007), (LABORSim: Leombruni and Richiardi, 2005) and links with macro-economic models: (Clauss and Schubert, 2007), (Wing, 2004).

One of the main driving forces behind the development of microsimulation models has been the phenomenon of the ageing population. This has come about, in part, because average life expectancy, throughout developed industrialised economies has

risen. At the same time, birth rates have fallen so there will be less people of working age to support the increasing number of retired people. This has implications for government finances because the demand for expenditure on pension benefits and healthcare for the elderly will be increasing at the same time that taxation revenues will be falling. Concern over the fiscal implications of population ageing has been one of the motivations behind the development of a range of models to project the future cost of pensions and healthcare. In the UK, one of the principal models used to project demand for pensions is Pensim2 (Hancock, Mallender and Pudney, 1992). This was developed in what is now the Department for Work and Pensions (DWP). It is designed to forecast the income of elderly people and the cost of pension provision under a range of policy options. In its current formulation, Pensim2 uses data from the Family Resources Survey (FRS), the British Household Panel Survey (BHPS) and the Lifetime Labour Market Base Data 2 (LLMBD2). These data are amalgamated to form a synthetic dataset that is projected to the year 2050. Pensim2 includes modules to project mortality, institutional care, education, disability, partnership, fertility, labour market status, job characteristics, earnings, national insurance contributions, housing, savings, pensions, taxes and benefits (Emmerson, Reed and Shephard, 2004). Pensim2 has helped to inform the debate over pension provision in the UK. One example of this was the Pensions Commission report (Pensions Commission, 2005) which led to the recommendations that spending on pensions should increase as a share of GDP and that the age of eligibility for a state pension should rise to between 67 and 69 by the year 2050 (Hills, 2006). As population ageing is occurring throughout the industrialised world, several other advanced industrialised countries

have developed models to assist pension policy. These include (APPSIM: Harding, 2007; Percival, 2007), (BRALAMMO: Zylberstajn et al., 2011), (MIDAS: Dekkers and Belloni, 2009), (MiMesis: Mikula et al., 2003) and (T-DYMM: Tedeschi, 2011).

Dynamic microsimulation modelling has also been applied extensively to inform policy on health. Spielauer (2007) and Rutter, Zaslavsky and Feur (2011) provide reviews of applications in this area. One UK model that focuses on health is CareSim (Hancock et al., 2007). This uses data from the Family Resources Survey (FRS) to project what individuals would have to pay for care in their own home or in a residential setting. It simulates mortality, income and capital as well as including assumptions on future care charges, taxes, benefits and GDP growth. Due to the uncertainties associated with assigning incomes to people as they enter retirement, the model is restricted to those aged over 65 and tracks this group, without adding new individuals, as they reach the age of retirement. CareSim is thus classified as a dynamic cohort model.

CareSim was used in conjunction with the cell based PSSRU (Personal Social Services Research Unit) model to investigate the effect of population ageing on long term funding for the NHS (Wanless, 2002) and later, the funding of social care (Dilnot, Warner and Williams, 2011). The latter study found that total expenditure on long-term care services, including health, social care and disability benefits used to fund care, is projected to rise from £20.6 billion (1.6% of GDP) in 2010 to £44.8 billion (2.3% of GDP) in 2030, in constant 2010 prices. Other health

microsimulations include (Long Term Care Model: Hancock, 2000), (LifePaths: Wolfson, 2000) and (HARDING: Harding, 1993).

A UK model that includes the simulation of both pensions and health is (SAGE: Simulating social policy in an AGEing society: Zaidi, 2007), developed at the London School of Economics in collaboration with Kings College London. Based on census, panel and cross-sectional survey data, it includes modules to simulate birth, death, education, marriage, divorce, labour force participation, earnings, health, retirement, disability and informal care (Spielauer, 2007). While Pensim2, CareSim and SAGE are all examples of UK based models, most other advanced industrialised countries have a functionally similar collection of microsimulation models, tailored to their own requirements and data sources (Zaidi, Harding and Williamson, 2009).

2.5 Microsimulation Models of Consumption and Expenditure

It is apparent from this overview of some of the major microsimulation models worldwide, that national governments are by far the most important sponsors and users. As a consequence, it is not surprising that these models are oriented towards answering the kind of questions that governments are interested in. Prominent among these are the issues of income from taxation, and government expenditure in the form of benefits, pensions and healthcare. Despite the importance of the sector in monetary terms, domestic consumption has been relatively lightly studied using microsimulation techniques. Nevertheless, some attempts have been made and prominent among these is EUROMOD which has been extended to include modules

to model expenditures (O'Donoghue, Baldini and Mantovani, 2004). This was done in order to study the re-distributional effects of indirect taxation. As indirect taxes are levied at the point where money is spent, the amount of tax each household pays is dependent on its spending pattern or how much of each commodity is consumed. This means that to predict the effect of indirect taxes on households, it is necessary to first develop a way of modelling their expenditure pattern. The first step in this was to model total consumption by estimating the parameters of an equation linking consumption C with income Y and a vector of socio-demographic characteristics X using data from respective European national household budget surveys.

$$\ln C = \alpha + \beta \ln Y + \gamma X + u$$

Where $\alpha, \beta,$ and γ are parameters to be estimated and u is an error term.

When this was done, the total consumption expenditure for each household could be imputed. Next, the budget share w of each good i could be obtained using the equation:

$$w_i = \alpha + \beta \ln C + \gamma (\ln C)^2 + \delta X$$

δ is a parameter to be estimated.

The knowledge of budget shares then allows the impact of indirect taxes to be calculated for each household.

Another model of this type is known as the Simulation Program for Indirect Taxation (SPIT: Symons and Walker, 1988). It was developed at the Institute for Fiscal Studies

(IFS) to predict the effects of indirect taxation on tax revenue, household expenditure and living standards. Its base dataset is derived from the Family Expenditure Survey and the categories of expenditure modelled are clothing, beer, wine, spirits, fuel, transport, services, petrol, tobacco, other goods, housing and durables – all items that attract indirect taxes. The core of this model consists of a demand system which takes into account the prices of each commodity group (Symons, 1991). SPIT has also been used to model demand for childcare (Baekgaard and Robinson, 1997) and household consumption behaviour (Symons and Warren, 1996). A related model also from the IFS is the Simulation Package for Energy Demand (SPEND: Baker, 1991). In the US, PRISM has been developed to model local residential telephone demand (Atherton et al., 1990). In Italy, Ando and Altimari (2004) developed a dynamic microsimulation model to study the effect of demographic change on income, savings and total consumption but consumption is not sub-divided into budget shares for particular items.

In many countries, rising life expectancy is leading to an ageing population where the number of people above retirement age is increasing relative to the size of the workforce. The fiscal implications of this transition have been investigated in a series of microsimulation models to project the effects of demographic change on healthcare and pension costs, some of which were mentioned above however the effect on consumption behaviour has been less well studied by microsimulation.

Beyond the microsimulation community however, there have been a number of

attempts, in several countries, to analyse the effect of an ageing population on household spending patterns. Lefebvre (2006) used a pseudo panel method to project spending for 10 composite goods in Belgium, finding that the share of aggregate expenditure on health, housing and leisure increases while equipment, clothing and transport decrease. Takeuchi (2009) investigated household expenditure for a range of goods in Japan by re-weighting observed spending patterns according to official population projections. He found that total consumption expenditure will increase to 1% above its 2005 value by 2010 and then declines to 7% below its 2005 baseline by 2030. There is some variation around this for different consumption categories. Spending on 'medical care' rises by 4% before returning to its original value by 2030. Spending on 'education' and 'transport' falls from its 2005 level and is below 90% of its original value before 2030. Luhrmann (2008) used a Quadratic Almost Ideal Demand System (QAIDS: Banks, Blundell and Lewbel, 1997) to project spending for 11 categories in the UK. She found that the aggregate budget share for 'fuel', 'household services' and 'personal goods' increases while spending on 'clothing', 'transport', 'leisure services' and 'food consumed outside the home' fell. In addition, she noted that the effect of changes in age and household type are moderate when compared with the effect of hypothetical increases in total expenditure or a redistribution of resources between generations.

One of the difficulties that previous work has encountered is a reliance on official projections. The parameters chosen by the various agencies are not necessarily the ones that are the most appropriate for modelling expenditure. Luhrmann and Lefebvre

for example, were both hampered by a lack of projections of future household size - a parameter which is expected to change as the population ages and also has a significant influence on spending patterns. Another limitation of previous work is that the use of equivalence classes means that some of the information on the heterogeneity of households is lost. All the methods used in the examples above (pseudo panel, re-weighting, and QAIDS) entailed some form of aggregation of cases into age-bands or cohorts.

In the following chapters, it will be shown in that the methods and tools developed in this thesis make it possible to avoid these difficulties, firstly by using demographic projections from Tyche (Chapter 4) to produce the independent variables for an expenditure system, it will be possible to project the most appropriate parameters for modelling household expenditure. Secondly, the expenditure system will be based on random assignment (Chapter 5) so that the individual cases are preserved and the full heterogeneity of households is represented throughout. Finally, in Chapter 6, the two components will be combined to develop a model to project the effect of population ageing on UK spending patterns.

2.6 Software for Microsimulation Modelling

One of the major tasks associated with microsimulation modelling is to develop the software. The preceding discussion gives some indication of the level of complexity involved in microsimulation modelling. This section reviews current methods and tools for its implementation.

Early microsimulation models were written in general-purpose programming languages such as C. While C is a powerful and flexible language, it requires the programmer to define all the variables, implement the code and control how the two interact. Throughout the computer industry, the general problem of increasing software complexity led to the development of object oriented languages. The advantage of these is that the code and data are defined in a single unit called an object which encapsulates some of the complexity of the software and facilitates modularisation of the program (Khor et al., 1995). The programmer can now isolate the code they are working on and be less concerned with the operation of other parts of the program. This modularisation also facilitates several programmers working on different parts of the program at the same time, as is common in large software engineering projects.

Object oriented languages like C++ are well suited to developing microsimulation models because there is a natural correspondence between the software object and the microsimulation unit. Objects contain all the information about the unit such as age, marital status and income, as well as what it can do such as form partnerships, give birth and eventually die. One of the difficulties that remains is that the model developer must write the code to manipulate the collection of objects. This includes adding and deleting objects, stepping through the list, sorting it into order and selecting a particular unit from the list. One way this can be done is by arranging the objects in what is known as a linked list. Here, each object is connected to the next by holding its memory address. This can be a daunting task if the modeller has no

previous experience of programming. Another area of difficulty is in developing the user interface. A high quality graphical interface that makes the model easy to use, requires specialist skills on the part of the developer. Quite often, microsimulation implementations make use of third party software to facilitate user interaction with the program. EUROMOD for example uses spreadsheets to enter model parameters and format the output into charts (Lietz and Mantovani, 2006).

2.6.1 UMDBS

One way to reduce the burden of programming is to hold the units in a database and access them through a database management system (DBMS). This is the method used in UMDBS (Universal Micro DataBase System: Sauerbier, 2002). Here, the microsimulation language MISTRAL (Micro Simulation Transformation and Analysis Language) is based on a cut down version of Pascal and includes a graphical user interface. However, this ingenious solution to the problem of software complexity does not seem to be used widely in microsimulation modelling.

2.6.2 LIAM

Life-cycle Income Analysis Model (LIAM) was originally created as part of a Ph.D. thesis (O'Donoghue, 2001b) in the form of a 'dynamic steady state cohort model' (O'Donoghue, 2001c: 196). It was written in C++ and its novelty is that it forms a flexible framework for the development of multi-cohort dynamic microsimulation models. This was built upon in the following years (Liegeois and Dekkers, 2011) and it has been used in developing dynamic microsimulation models in several European

countries (O'Donoghue, Lennon and Hynes, 2009). The latest version, LIAM2, is written in Python but the user accesses it through a YAML markup language interface. LIAM2 is considered more fully in Chapter 4.

2.6.3 Modgen

Another attempt to alleviate the problem of software complexity is Modgen produced by Statistics Canada. This general-purpose microsimulation language hides the underlying mechanisms for controlling the units. It also provides a visual interface, documentation and a range of tools for tabulating the output. This frees the modeller from a good deal of low-level programming and allows them to concentrate on setting up the parameters, units and events. Modgen operates as a C++ preprocessor. This means that models created using Modgen are really implemented in C++ but the model developer does not program in that language. The developer creates the model using the Modgen language and the preprocessor converts this into C++ which is then compiled and executed by the computer. Microsoft Visual C++ Developer's Studio is required in addition to Modgen in order to develop Modgen models but once compiled, they can be run as an executable file on any Windows computer.

Modgen has been used by Statistics Canada to develop a wide range of microsimulation models (Statistics Canada, 2011). These include Pohem (Population Health Model), IDMM (Infectious Disease Microsimulation Model), CVMM (Child Vaccination Microsimulation Model) and LifePaths which is a general-purpose dynamic microsimulation model. In addition to a set of continuous-time dynamic

microsimulation models, which model demographic change, health and pension sustainability, the versatility of Modgen has allowed it to be used to develop a model of predator prey interaction and a simulation of an experiment to study the life-cycle of the tsetse fly. Modgen is a powerful language that considerably reduces the complexity of software development. Despite this, as a superset of C++, it still requires the modeller to 'have some understanding of structured programming principles' (Statistics Canada 2008; 5).

2.7 Challenges for Microsimulation Modelling

The development of a dynamic microsimulation model is an ambitious undertaking. It attempts to represent a society of individuals and their decisions regarding partnership formation and dissolution, having children and leaving home. On top of this there is the substantive purpose of the model which may be to represent the pension system, health trajectories or taxation system in one or more countries. As mentioned above, this requires large quantities of high-quality data and significant computational resources, both in terms of hardware and the development of the software. The enterprise can also be costly and time consuming where, for example, the development costs for MINT, CBOLT and POLISIM were over 6 million US dollars. The funding for SAGE amounted to £800,000 per year for five years while APPSIM cost a total of 1.7 million Australian dollars over five years (Harding, 2007). The development of a dynamic microsimulation model is therefore a long term project which may not sit easily within the usual university funding cycle (Li and O'Donoghue, 2013).

The complex nature of microsimulation has meant that ease of use presents a significant challenge (Merz, 1993), (Harding, 2007). This is the case for the end user who is interested in running simulations and interpreting the results. It also applies to developers of microsimulation models who require the most advanced software engineering tools. Unfortunately, the aim of developing an advanced graphical interface for the end user places an additional burden on the developer. Schofield (1995) noted that the design and development of a Windows-style user interface for a static microsimulation may take about a quarter of the development time for the whole model. A similar ratio can be anticipated for dynamic models.

The issue of validation has also been problematic in dynamic microsimulation modelling (Merz, 1993). Morrison (2008) describes a multi-level validation method comprising: data, algorithm, module and impact validation but Li and O'Donoghue (2013) note that, compared to the amount of work done in other areas, limited effort has been placed on validation and there is no international consensus on validation procedures. One effective way to validate the results from a dynamic microsimulation model would be to compare its projections with observed data. However as Winder (2000) noted, microsimulation models usually fail to simulate known time-series data. The reason for this was discussed above and is due to a reliance on past data to develop the model so that even if the model is perfectly specified and sampling error is ignored, model projections tend to drift away from observed data. Li and O'Donoghue (2013) attribute a perceived failure of early microsimulation models to aspiring for unachievably high expectations of prediction and accept that human

behaviour is so complex that dynamic microsimulation models cannot hope to make highly accurate projections.

Microsimulation models are often developed within a governmental or policy environment where dissemination has a lower priority than it does in academia (Li and O'Donoghue, 2013). As a result, knowledge of methods is often only available through technical reports or model documentation. This leads to a situation where developers have reinvented techniques several times and there is a lack of accessible material for training new developers.

Apart from partnership formation, interaction between microsimulation units has not been implemented extensively (Li and O'Donoghue, 2013) and the representation of behaviour has been quite limited (Wolfson, 2009). When behaviour is represented, it is usually done by means of equations and this is not conducive to modelling interactions between units. Also, socio-economic behaviour at the micro-level is highly complex and appropriate micro theory and data are limited (Bacon and Pennec, 2009).

This section has identified some of the most significant difficulties in developing dynamic microsimulation models. The proposal of using an agent-based modelling toolkit to reduce the burden of developing the software directly approaches only one of these which is to make the process easier. However, it will be argued in section 2.9 below that there will be indirect benefits in several other areas that arise from this,

such as reducing costs, improving usability, implementing behavioural interaction and making the field more accessible to new modellers.

2.8 Agent-based Modelling

Although they were developed independently from microsimulation, agent-based models (ABM) also work at the level of individual units. This section describes the origins of ABM and the types currently in use. It uses the results from previous reviews to assess the suitability of ABM toolkits for microsimulation modelling. This leads on to the justification for selecting a script-based ABM platform called NetLogo as a way to reduce the burden of developing the software in microsimulation modelling.

2.8.1 Origins of ABM

The origins of agent-based modelling arise from the idea of a self-replicating machine, now known as the von Neumann machine (von Neumann, 1966) (Iltanen, 2012). This itself was derived from a kind of Universal Computing Machine or Turing machine (Turing, 1937). The Turing Machine was developed to investigate the theoretical limits of computation and was intended to represent the simplest possible machine that was capable of performing computations. It consists of a tape of indefinite length that holds a set of symbols. A read-write head moves over the tape and can change the symbols according to the internal state of the head.

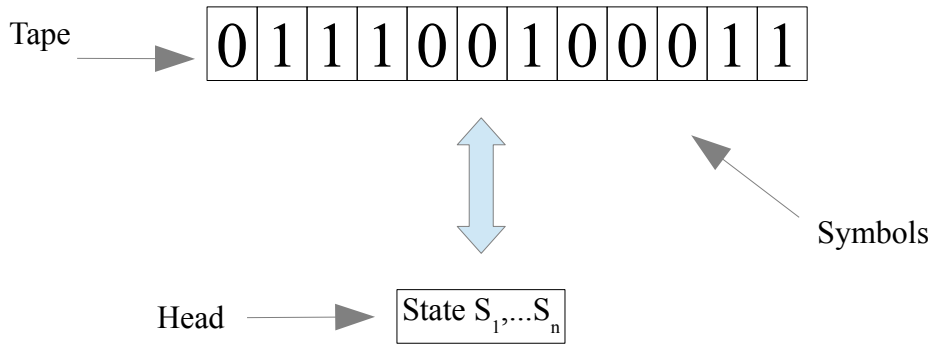


Figure 2: Schematic Representation of a Turing Machine

For any combination of state and symbol read from the tape there is only one possible next state. This can be represented in a state transition diagram.

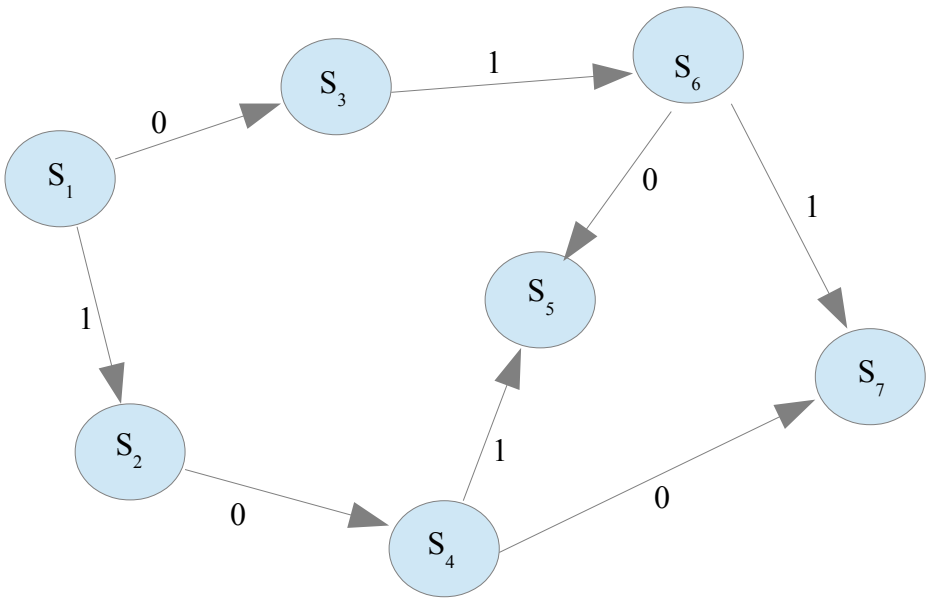


Figure 3: Turing Machine State Transition Diagram

In figure 3, if the head is in state 1 and a 1 is read from the tape then the head will transition to state 2. If a 0 had been read then the next state would have been state 3.

In this way, the Turing Machine can be said to be deterministic because there is at most one possible transition corresponding to each combination of state and input symbol.

In the sense that it reads input data, has an internal state which controls its operation and can modify its environment by producing an output, the Turing Machine corresponds to what we now would call a cellular automaton (Iltanen, 2012). The next major innovation was to combine several machines into a formation where they could interact such that the output from one became the input for the next. This was done in the 1970s when Conway implemented the 'game of life' (Gardner, 1970), which became widely known to computer enthusiasts at the time. The basis of this game was a grid of cells, which could be 'alive' or 'dead' represented by being black or white on the computer screen. Each cell was, in a sense, aware of its environment and altered its state according to simple rules. If a cell is dead and three neighbouring cells are alive then it becomes alive. If more than three neighbouring cells are alive, the cell dies of overcrowding. If less than two neighbouring cells are alive, the cell dies of isolation.

These cells had a fixed location. One of the first examples of a model implementing mobile agents appeared in a simulation of the flocking behaviour of birds called Boids (Reynolds, 2008). The simulated birds operated under rules for separation, steer to avoid crowding local flockmates; alignment, steer towards the average heading of local flockmates; and cohesion, steer to move toward the average position

of local flockmates. By varying the rules, it became possible to simulate a range of other creatures such as herding cattle and schools of fish as well as human behaviour like co-operating in a social network or forming alliances between nations (Axelrod, 1995).

2.8.2 Behavioural Representation in ABM

Many agent-based models use highly abstract representations of decision making behaviour. A well known example of this type is the iterated prisoner's dilemma (Axelrod, 1984) in which the agents decide to cooperate with their partner in crime or defect and inform the authorities, on the basis of a predefined, although in some cases, adaptive strategy. Another example is the Schelling model of residential segregation (Schelling, 1969). This simulation is based on the assertion that people have a preference for living in proximity with others who are similar to themselves. If the proportion of dissimilar people in an area rises above a certain threshold, people will move to a more preferable area. The surprising outcome of this model is that even if people are prepared to live in an area where there is a significant majority of other types, the population as a whole organises itself into highly segregated regions with little mixing. The phenomenon of large scale patterns emerging out of local interactions is a recurring theme in agent-based modelling. Emergence is defined as the formation of novel and coherent structures, patterns and properties during the process of self-organisation in complex systems (Goldstein, 1999). An example of this is the appearance of stationary waves in dense traffic (shock-waves), which can be reproduced in many of the traffic microsimulation models mentioned above.

Epstein and Axtell (1996) use the concept of emergence to propose a new theoretical and methodological perspective on the formation of human societies.

While the proponents of abstract models, such as Epstein and Axtell claim that their method uncovers profound and general features of the social world, other agent-based models are underpinned by a more specific, or realistic, representation of the mental processes of the simulated units. One common approach is: beliefs, desires and intentions (BDI: Bratman, 1987). In this scheme, the three components or attitudes model the informational, motivational and deliberative states of the agent which makes rational choices to achieve adequate or optimal performance when deliberation is subject to resource bounds (Rao and Georgeff, 1991). Several varieties of cognitive model exist such as intention theory (Cohen and Levesque, 1990) and PECS (physical conditions, emotional states, cognitive capabilities and social status) (Schmidt, 2000). This has been used in an agent-based model of crime (Malleson, 2012). A further development in agent complexity is to found behaviour on sociological theory such as Trigg, Bertie and Himmelweit (2008) who use Bourdieu's theory of practice (Bourdieu, 1977, 1984) to model cultural diffusion while Henrickson (2002) uses it to model educational choice. Edmonds and Chattoe-Brown (2005) investigate the use of social network analysis (Scott and Carrington, 2011) in simulations. The online Journal of Artificial Societies and Social Simulation (JASSS) http://jasss.soc.surrey.ac.uk/JASSS.html provides a source of information on current developments in ABM.

2.8.3 ABM and Economics

The application of ABM to economics is often known as agent-based computational economics (ACE). In a survey of this field, Tesfatsion (2002) notes that ACE operates on the principle that economies are complex adaptive systems consisting of large numbers of adaptive agents involved in parallel local interactions. Significant areas within ACE include: market processes, the development of norms, the formation of economic networks, modelling organisations and automated markets. ABM has also been used to model the behaviour of consumers. On the supply side, Heppenstall, Evans and Birkin (2006) developed a spatial model to examine price setting at petrol stations where decisions were based on local competition and profit earned in the previous period. In consumer demand modelling, Venables and Bilge (1998) simulate how customers more around a supermarket as they select products from the shelves. Collings et al. (1999) model the rate, extent and order of product diffusion in the telecommunications market by representing the interaction of consumers within a social network. Although there are several ACE models that represent an economy or market as a whole, as in Westerhoff and Franke (2012) or focus on a particular segment or type of good, such as scarce or positional goods (Bernardino and Araujo, 2009) or lighting (Afman et al., 2010) there do not seem to be any examples that model a complete household budget set, which is the aim of this thesis. Tesfatsion and Judd (2006) provide a comprehensive review of developments in this field.

2.8.4 ABM Software

Despite the wide range of application areas and the variety of internal representations,

the underlying similarity in the structure of the agents made it apparent that much of the software could be implemented in the form of a toolkit. This leaves the modeller free to concentrate on specifying the rules for the agents to follow. Swarm was the first widely available ABM toolkit. This provided a set of classes to model a hierarchy of agent collections or swarms. It also included the code for an 'observer' to monitor agent behaviour as well as tools for the user interface. Swarm was originally written in Objective C and subsequently, a version in Java has become available which uses Java classes to call the underlying Objective C routines. Since the late 1990s, a range of 'Swarm clones' has been developed written entirely in Java. These include Repast, MASON and JAS (Java Agent Simulator). Recently, ABM toolkits that enable simulations to be created using scripting languages have become available. Traditional programming languages use what is known as a compiler to convert the source code, written by the programmer, into object code, which is executed by the computer. Scripting languages provide pre-compiled blocks of code which are run by the computer directly so that the programmer does not compile the code they write. This can simplify the development process. Repast allows models to be constructed using Java classes or a versatile scripting language known as Python. Others such as NetLogo (Wilensky, 1999) provide their own scripting language. The following list provides a summary of some common examples of ABM platforms. Nikolai and Madey (2009) provide a comprehensive review.

Ascape is a Java based Swarm like agent based simulator. Quite complex models can be created without any programming but for advanced work, Java

skills are required.

Breve is a 3D Simulation framework for artificial life and decentralised systems. It uses an Objective C like object orientated scripting language called Steve or alternatively, a scripting language known as Python which is also used in other simulators such as Repast. Breve supports animation and is available through GNU public licence (GPL) source-code. It is designed to be accessible to modellers with no or limited programming experience.

Cormas is designed for applications in the area of ecological and natural resource interactions but its framework of classes, observer and grid space gives it a wide area of applicability. Based on Visual Works programming environment, applications are developed by modifying generic SmallTalk classes.

Java Agent-based Simulator (JAS) was developed to make the functionality of Swarm available in Java. A model is implemented by modifying the jas.engine.SimModel. This provides the skeleton code for creating a collection of agents and organising them in a schedule of events. It also contains classes to implement neural networks and genetic algorithms. In addition to the built in packages that control the model and agents, a vast library of packages provide functions such as plots, graphics and links to external spreadsheets and databases. The graphical user interface supports the running of the model.

JAS is licensed under GNU Lesser General Public Licence.

MadKit provides a platform for developing multi-agent simulations. The framework is based on the concepts of observer, group and agent. The core of the system is written in Java but the behaviour of agents can be written in scripting languages like Python or can be implemented in Java. MadKit is licensed under GNU General Public Licence. It is free for non-commercial applications but restrictions apply for commercial use.

Magsy is an expert systems production language, which means that the agents are given a set of rules which control their action. The language used to program the agents is derived from OPS5 but has been extended to better facilitate agent-based modelling.

Mason provides Java libraries for developing multi-agent simulations. Swarm like but its design priority is execution speed.

Mimose provides a modelling language and experimental framework that is tailored for the social sciences. This involves non-linear and multi-level processes. It aims to relieve the model developer from as much programming and implementation details as possible.

NetLogo has been used in the area of modelling natural and social

phenomena. Written in Java, it comprises a visual grid which represents the agent's world and also includes an observer. The agents can be based on static regions called 'patches' or mobile 'turtles'. The origin of NetLogo in education, combined with good documentation, ease of use and library of models makes this one of the more accessible platforms for agent-based modelling.

StarLogo is a simulation platform designed with ease of use as a priority. It has a wide range of application areas such as modelling physical, biological and social processes. StarLogo is written in JAVA and later developed into NetLogo. There is now an open source version called OpenStarLogo. Another new variation of StarLogo is called StarLogoTNG. This makes modelling easier by providing syntax elements of program code on a menu. It is tailored towards graphics and game development.

Sugarscape is a social science simulator where agents move around a grid and consume simulated sugar as food. It has been used to simulate emergent behaviour and an economic system.

Swarm is the original agent based simulator. Originally written in Objective C, it now has a Java version that uses Java classes to call the underlying C code. It requires good Java skill to use and appears to have been superseded by numerous clones.

2.8.4.1 Evaluation of Agent-based Modelling Toolkits

Tobias and Hoffman (2004) make a comparison between four ABM toolkits with the aim of finding the most suitable platform for social scientific agent-based computer simulation. These are Repast, Swarm, Quicksilver and VSEit. They describe applications-oriented social simulation as being characterised by large numbers of complex agents that contain several social scientific theories. The simulations are grounded in empirical data where social networks are modelled in a way that is functionally equivalent to the real social world.

While it would be possible to program these simulations using a native language such as Java, the principle advantage of using a toolkit is the potential for saving time and effort on the part of social scientists. The evaluation begins with the formation of a list of potential candidates that were documented on the internet. This 'long list' was reduced to four by the application of a set of selection criteria. These were firstly that the framework allows simulation that is based on scientific theory and empirical data on complex micro-level objects. Secondly, the software should be freely available to facilitate collaboration with other social scientific institutions. Finally, the software must use Java as the implementation language as it is a widely used and powerful language that the authors believe may develop into a standard.

The four platforms that remain after the application of these criteria are then evaluated against a comprehensive set of requirements using a points system. In their tests, Repast scored the highest with 80 points. Next was Swarm with 71 followed by

Quicksilver with 61 and VSEit with 57. A modification to this scoring system was to apply a weight to each criteria depending on the effort spared for the user by the feature, the necessity of the function and whether the user can improve the function. Under the weighted system, the scores were Repast 311, Swarm 266, Quicksilver 241 and VSEit 227.

A different set of toolkits was reviewed by Railsback, Lytinen and Jackson (2007) which were: Objective C Swarm, Java Swarm, MASON, NetLogo and Repast. These platforms were chosen because in the author's experience, they were some of the most widely used at the time. The review takes place by developing an agent-based model that progressively becomes more complex and assessing whether and how well, each platform facilitates the process. The simplest model consists of 100 agents called bugs that can move around on a grid. The most complex model has a variable population of bugs that consume food from the cell at their current location. They grow, reproduce and die as well as being predated by a second species that hunts the bugs.

The authors noted that all the platforms reviewed could implement all levels of the test model. From the point of view of a novice user, NetLogo was found to be by far the easiest platform to use and it was also the best documented. However, at the time of writing, in 2007, it had some shortcomings associated with determining the order of execution of the agents. The current version of NetLogo, version 5, is open source so it is now possible to resolve these issues. The problems associated with the Swarm

type ‘framework and library’ platforms, which include Repast and MASON, mainly centre around them being difficult to use and this is exacerbated by poor and incomplete documentation.

The development of the series of models also allowed a comparison of the execution speed for each platform. The design priority for MASON is speed and this was the fastest platform for all versions except model 1. Repast was 54% slower than MASON in the simplest model but for the most complex models it achieved almost the same speed. NetLogo was nearly as fast as MASON in model 1 but became 3 to 5 times slower in the later, more complex models. Both versions of Swarm were faster than MASON in the simplest model, with Objective C Swarm being much the fastest. However, in the later models, this platform was 7 to 14 times slower than MASON and Java Swarm was the slowest with an execution time 19 to 31 times that of MASON.

ABM toolkits are beginning to be used for developing microsimulation models. One example is LaborSim (Leombruni and Richiardi, 2005) which has been used to study the effect of recent changes in retirement eligibility criteria in Italy. Another is IFSIM (Baroni, Zamac and Oberg, 2009). This is an ABM which forecasts demographic change taking place in the Swedish population and includes modules covering fertility, mortality, matching and leaving home. There are also modules that simulate the labour market, tax, benefit and pension systems as well as human capital and education. This allows recursive modelling of the effect of demographic changes on

the economy as well as the reaction of economic changes on demography through fertility rates.

Both LaborSim and IFSIM were developed using JAS (Java Agent-based Simulation library). This is one of the 'framework and library' types of environment. It is based on Java and so requires knowledge of object oriented programming as well as how to use the toolkit. The reason for considering using an ABM toolkit for developing the microsimulation model is to reduce the burden of writing the software. Out of the platforms reviewed above, NetLogo is the most promising in this regard. While it is the highest level, in terms of its language constructs, it was possible to develop all the test models in Railsback, Lytinen and Jackson's series. NetLogo is well documented, has a wide user base and its operating speed is mid-range among the ABM toolkits reviewed. It is the easiest to use and has an advanced graphical user interface. The next section examines NetLogo's features and suitability as a tool for microsimulation modelling in more detail.

2.8.4.2 Review of NetLogo

NetLogo is one of the Logo family of platforms which utilise a high-level scripting language rather than programming Java classes. On its home page, NetLogo is described as 'a cross-platform multi-agent programmable modelling environment.' It was created by Uri Wilensky in 1999 and is under continuous development at the Centre for Connected Learning at Northwestern University. It is written in Scala and Java with the aim of facilitating the development of cellular automata and agent-

based models. NetLogo provides a world consisting of a grid of 'patches', which are static. On this world, agents called 'turtles' can move around and an 'observer' keeps track of the whole simulation. On starting NetLogo, the user is presented with a graphical interface as shown in Figure 4. The main features are contained in a window where the simulation is represented. This consists of a 'world grid', which can be updated to view the progress of the simulation.

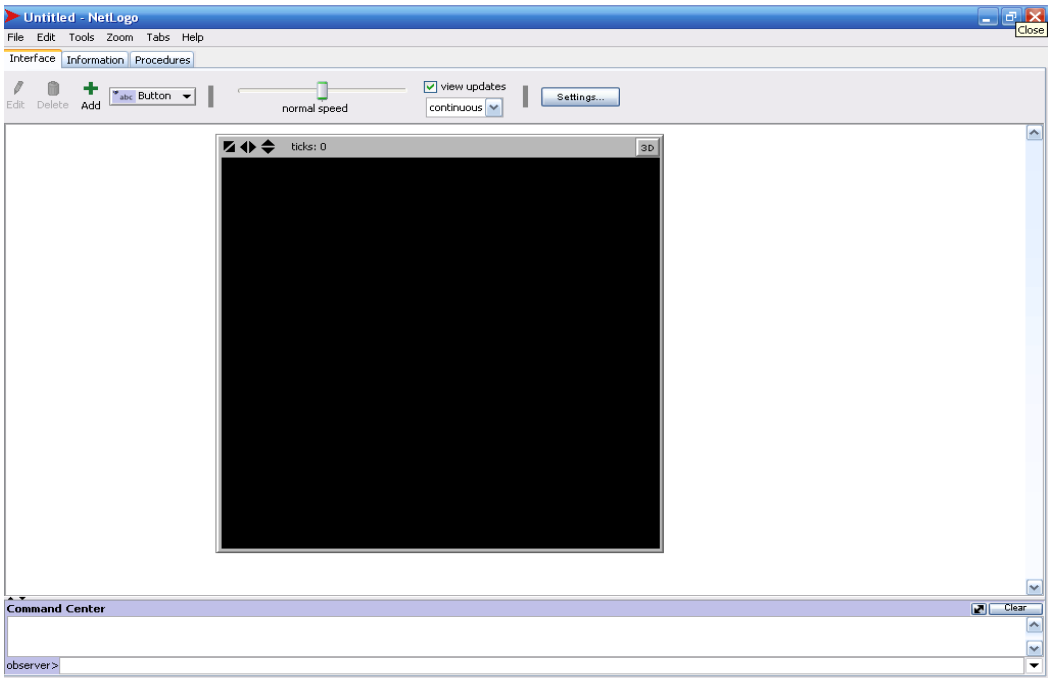


Figure 4: NetLogo Interface

A range of GUI elements such as buttons, sliders and plots can be added to track the progress of variables of interest. There is a tab named 'information' which provides an area where the documentation for the model can be stored. The 'procedures' tab provides access to the built-in editor where the simulator code is written and tested. NetLogo can be downloaded free of charge and the licence is 'unrestricted, including

commercial use' (NetLogo, 2010). An application programmer's interface (API) is provided for controlling NetLogo from external Java code if desired and 'extensions' can be created to allow the programming of additional functionality.

NetLogo is well documented. The website provides detailed information for using and programming the toolkit in printable and browser readable form. Tutorials are available and a range of example models are described in a 'models library' which can be accessed from the menus once NetLogo has been installed. An active user group exists to provide support in developing models.

The language itself is based on a set of high-level commands. Turtles can be created with the 'create' (CRT) command, for example CRT 6 creates six new agents. On creation, each agent is automatically assigned a unique identifier called the 'who' number. This can be used subsequently to address a particular agent and request that it performs some action. The command 'ASK TURTLES WITH [WHO = 25] [SET COLOR "GREEN"]' would make turtle 25 turn green. Agents can be manipulated with various types of move command. MOVE FORWARD 1 causes the current agent to advance to the adjacent patch. RIGHT 5 would change its course 5 degrees to the right. In addition to commands, there are 'reporters' that return a value. For example, SHOW COUNT TURTLES-ON NEIGHBORS would display how many agents were on the patches that are adjacent to the current patch. The state variables for the agents are declared within a block of code that begins with TURTLES-OWN. The state of the patches is set up in the PATCHES-OWN block. The concept of 'breeds' facilitates

the use of heterogeneous types of agents using the command BREED [breedname] which sets up a new type of agent entitled whatever 'breedname' is specified. Each breed has its own set of state variables.

From this review, it seems that NetLogo is a powerful yet relatively easy to use tool for agent-based modelling. The GUI enables rapid prototyping and development. The command language provides the full functionality of a programming language while the high-level commands, specific to modelling, make it easy to learn. NetLogo is open-source so it is possible to find out exactly how the primitives are implemented if necessary. The 'patch' 'turtle' format meshes very well with the requirement to simulate households and individuals in microsimulation modelling. The static patches can be used to represent households while the mobile turtles represent individuals who can move between households as they progress through their life-course. These advantages seem to make NetLogo a promising candidate for a microsimulation modelling platform and Chapter 4 tests this idea by developing and validating of a dynamic microsimulation model using NetLogo.

2.9 Microsimulation and Agent-Based Modelling

This review has introduced both microsimulation and ABM, which provides an opportunity to compare their essential features, strengths and weaknesses. As was noted above, ABM originated from the study of mathematics and artificial intelligence. It arose from the Turing Machine where transitions between states were deterministic. For any given combination of machine state and input read from the

tape, there is only one possible next state. ABM is not aimed at a particular application area but is well suited to abstract or theory-based models with interacting entities. In many cases, the aim is to study emergent phenomena that arise from the interaction between micro-level units. Models have been created within the diverse areas of physics, climatology and the social sciences. Much of the versatility of ABM applications can be traced to the fact that the behaviour of the agents is formulated as a set of rules. This means that any system that can be described in terms of rules can, in principle, be implemented as an ABM. One difficulty associated with this approach is that it is necessary to specify all the rules for the agents follow. This is especially problematic when modelling social systems where the rules are complex, ill-defined, change over time or simply not known. Another weakness associated with ABM is that few implementations are supported by empirical data (Gilbert and Troitzsch, 2005), (Hassan et al., 2008).

By contrast, microsimulation modelling was conceived to solve a particular problem associated with aggregated measures. Typical applications are the modelling of demographic change and tax-benefit systems, often by government agencies to answer specific policy questions. Changes in the state of microsimulation units are often implemented in a set of transition probabilities. The approach is supported heavily by the use of observed data and statistical analysis but the interaction between units is typically restricted to partnership formation and behavioural modelling has been quite limited (Wolfson, 2009), (Davidsson, 2001). The essential differences between ABM and microsimulation are highlighted in the following table. These are

generalisations and there may be exceptions in particular cases.

	Microsimulation	**Agent-based Modelling**
Focused on a Specific Policy Question	Often	Not usually
Behavioural Representation	None or rational agent	Diverse - from abstract rules to sociological theory
Emergence	Usually absent	Often represented
Level of abstraction	Low, uses empirical data	High, uses abstract concepts
Use of theory	Some economic theory	Wide range of theories
Interaction	Limited to partnerships	Varied, complex
Learning	None	Possible
Communication	None	Possible
Decision Rules	Mostly probabilistic	Usually deterministic
Based on empirical data	Usually	Rarely
Cost and Complexity	High	Low to medium

Table 1: Summary of Microsimulation and Agent-based Modelling Characteristics

These differences in the type of question each method is intended to answer and also the way of going about answering them makes ABM and microsimulation seem to be quite distinct disciplines. Chattoe-Brown (2009) doubts whether the two can be brought together because of an inherent conflict between the behavioural perspective of ABM and the focus on attributes that is a feature of microsimulation. In this context, it is difficult to see how the advantages of both approaches can be combined so that microsimulation can benefit from the more theoretical approach of ABM and how ABM can be given a more sound empirical foundation by incorporating agents based on real data.

Nevertheless, there have been some attempts to bring together some features of the two approaches. Hassan et al. (2008) advocate the use of empirical data to set the initial conditions of an ABM. A model combining spatial and dynamic microsimulation was augmented with an ABM component to model population change in Leeds (Wu, Birkin and Rees, 2008, 2010). Mahdavi et al. (2007) propose a multi-layer method of modelling residential segregation. Here, an agent-based modelling layer simulates the behaviour of individuals while a microsimulation layer represents the transitions between states for larger units such as households and businesses. The layers interact so that both upward and downward causation can be modelled. Also, in this thesis, the combined model of Chapter 6 uses a standard dynamic microsimulation model based on transition probabilities, with an expenditure system based on random assignment that operates using a simple interaction based copying mechanism. These two components operate in a way reminiscent to the separate layers as proposed by Mahdavi et al. In this way, the model might be characterised as an implementation of a multi-layered hybrid agent-based microsimulation.

These links between ABM and microsimulation are made by combining elements of the two approaches at different points in the same model. However, it can also be argued that the approaches themselves are less different than it might appear from an account of the applications to which they have been applied. While interaction is often given a marginal role in microsimulation, Orcutt was clear that his units were intended to interact when he first introduced them as ‘various sorts of interacting

units' (Orcutt, 1957: 117). Also, the extensive use of probabilistic models to represent behaviour in microsimulation modelling can be explained by the lack of knowledge in the domain in which the model is applied. When faced with an absence of knowledge of a particular social process, microsimulation modellers often turn to statistical methods to determine the probabilities of various outcomes. However, this is a limitation imposed by lack of knowledge of the thing being modelled rather than being a feature of microsimulation itself.

Agent-based models can also incorporate probabilistic behaviour by the use of a random number generator that is an integral part of most computer language implementations. This addition converts the agent into a probabilistic Turing Machine where some transitions are random choices among finitely many alternatives (Black, 1999). In this way, the agent acquires the same structure as a microsimulation unit with inputs, outputs and a set of operating characteristics which convert inputs to outputs on the basis of deterministic or probabilistic means. In the light of these considerations, it seems to be the adaptation of the technology to particular problem areas that separates microsimulation from ABM rather than a fundamental difference in the technologies themselves.

2.10 Evaluation of the Project

Section 2.7 above discussed some of the difficulties in developing dynamic microsimulation models which can be summarised in the following list:

- complexity of software
- cost and development time
- poor usability
- difficult to validate
- limited behavioural representation
- poor accessibility
- lack of predictive power

In terms of its contribution to microsimulation, the use of NetLogo is intended to make the software easier to develop and so will reduce the complexity of developing dynamic microsimulation models. Since this problem can be seen as the source or a contributing factor to many of the other difficulties listed above, it will have an indirect impact on several others.

NetLogo provides a set of high-level commands with which to manipulate the collection of agents. As a result, it can reasonably be expected to have the potential to reduce development time and cost because the developer will not have to code and debug the low-level elements which are encapsulated within the scripting commands. The GUI, which is built into NetLogo represents the current standard in interface design that is widespread in many areas but is not commonly implemented in dynamic microsimulation models. This will contribute to improving the usability of dynamic microsimulation modelling. The correspondence between agent-based modelling and microsimulation was discussed above. The use of NetLogo makes it

natural to import the flexibility of behavioural interaction into the microsimulation model. This is utilised in implementing the copying mechanism for allocating budget sets to households and so provides a prototype for adding behaviour to microsimulation models. The high-level scripting language, GUI and graphs that update as the program is running provide an accessible example of a dynamic microsimulation model which has the potential to provide a resource for introducing the next generation of microsimulation modellers to this field.

Other items on the list are ongoing problems that present difficulties for microsimulation modelling in general but are not addressed here. The thesis does not contribute to making microsimulation models better at prediction. It was observed above that overambitious aims for prediction may have damaged the credibility of microsimulation modelling in some quarters. This thesis advocates the use a combination of dynamic microsimulation and alignment to obtain the advantages of both approaches. The thesis does not add to the ability to validate microsimulation models. While some progress has been made in the validation of static models, such as Pudney and Sutherland (1994), further work is needed in validating the more complex dynamic microsimulation models. However, the thesis makes use of Morrison's scheme, which may be considered to be the most sophisticated approach currently available and investigates the effect of stochastic variation which is introduced by the random assignment component.

2.11 Conclusion

This chapter has provided a review of current technologies for modelling at the micro-level. It began with an overview of modelling and simulation before embarking on a more detailed study of microsimulation methods. This showed that despite the wide array of approaches available, the construction of a dynamic microsimulation model is still a complex and burdensome exercise which may inhibit the more widespread use of this technology. Other weaknesses were found to include: high development costs, poor usability, difficulty of validation, limited behavioural representation, poor accessibility and a lack of predictive power. The apparent similarity between the 'interacting units' of microsimulation with those found in agent-based modelling led to the proposal that an ABM toolkit may be useful in simplifying the development of a microsimulation model and have some indirect benefits in addressing some of the other weaknesses. ABM platforms have been used in microsimulation modelling before but these have been of the ‘framework and library’ type that incorporate a general-purpose programming language. There has been no attempt to apply a script based ABM toolkit. Later, in Chapter 4, this research will test the suitability of what may be the simpler to use, NetLogo, as a dynamic microsimulation modelling platform, in terms of functionality, processing speed and usability.

Chapter 3: Expenditure Modelling

3.1 Introduction

This chapter is concerned with the question of how household expenditure can be modelled at the micro-level. It reviews two quantitative approaches to modelling consumption. These are time-series analysis and demand systems. It is argued that time-series methods are not well suited to representing expenditure at a disaggregated level because it is not possible to decompose the time-series into sub-categories. Also that demand systems present a number of difficulties in dealing with the heterogeneity of consumers, which is an issue of central importance in microsimulation modelling. Random assignment, as introduced in Chapter 1, is proposed as an alternative that is compatible with modelling at the micro-level of individuals or households while retaining all the information on the heterogeneity of households.

Time series methods are reviewed first. Then the development of demand systems is described, which leads to the identification of some of the difficulties encountered when applying this approach at the micro-level. Finally, random assignment is proposed as a way of overcoming the limitations of current approaches in the context of microsimulation modelling.

3.2 Time-series Models

This approach relies on a set of observations of the parameter of interest, which have

been collected at regular intervals over a period of time. If the process generating the data is not completely random, there may be detectable correlations between pairs of observations collected with a constant separation or lag. An autoregressive (AR) model of the time-series can be constructed which is a linear regression of the current value of the series against one or more previous values (NIST/SEMATECH, 2003). Its general form is given by:

$$X_t = \delta + \phi_1 X_{t-1} + \phi X_{t-2} + ... + \phi_p X_{t-p} + A_t$$

where X_t is the time series at time t, A_t is white noise (a random variable which has a constant mean average and is not autocorrelated) and

$$\delta = (1 - \sum_p \phi_p)\mu$$

where μ is the mean value of the series.

The parameters ϕ_1 . . . ϕ_p can be estimated using standard statistical software.

Another method for analysing time-series data is the moving average (MA) model. This approach effectively regresses the current value of the series against the white noise of one or more previous values. Here,

$$X_t = \mu + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - ... - \theta_q A_{t-q}$$

where X_t is the time-series at time t

μ is the mean of the series

A_{t-q} represents the white noise

ϕ_1 . . . ϕ_q represent the parameters of the model

Box and Jenkins (1970) combined these two approaches in the Autoregressive

Moving Average model (ARMA). Many time-series approaches including ARMA assume stationarity where the mean, variance and autocorrelation structure of the series do not change over time. When an observed time-series is differenced by subtracting adjacent pairs of data points to achieve stationarity, the model is known as an ARIMA (Autoregressive Integrated Moving Average) model. Montgomery, Jennings and Kulahci (2008) provide a comprehensive introduction to time-series methods.

One of the most important applications for time-series methods is forecasting (Clements and Hendry, 2002). In the particular area of stock market trading, technical analysts believe that each point in the time-series reflects the knowledge of all market participants (One Asia Capital, 2014) and so its variations over time, such as trends, can be used to predict future prices. When applied to predicting household demand, time-series methods often work at the aggregated level such as Mohtashami and Salami (2009) who compare a range of time-series methods of forecasting demand for basic food types. McLoughlin, Duffy and Conlon (2013) compare various approaches for modelling electricity demand at the household level using data obtained from smart meters.

It is also possible to combine a number of time-series together in Vector Auto Regression (VAR) to take into account the relationship or influence of a number of interacting variables. However, it is unlikely to be possible to use time-series methods to model a complex budget set at the household level because the longitudinal dataset

that would be required, is rarely available. The difficulty of using time-series methods at a disaggregated level means this approach is more commonly applied to macroeconomic modelling.

3.3 Regression Methods

Simple linear regression is a widely used statistical technique for representing the relationship between two parameters or variables (Montgomery, Peck and Vining, 2001). If the variables are plotted on a graph, the regression equation is represented by a straight line, drawn such that the variation of the data points above and below the line is minimised. This 'line of best fit' is represented by the equation:

$$Y = b_0 + b_1 X$$

Y is known as the dependent variable because its value is thought to depend on the value of X which is the independent variable. b_0 and b_1 are constants that can be obtained using standard statistical software. For this approach to be valid, there should be a linear relationship between the dependent and independent variables, the independent variables should not be highly correlated with each other, the errors should be normally distributed with a mean of zero and should have constant variance. However, the method is robust to small departures from these conditions. Multiple linear regression is an extension of simple linear regression to include more than one independent variable (ibid). In this case, the equation takes the form:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$$

and so on for as many parameters as the model requires. To construct a regression model for household demand, the variable to be predicted, for example monthly

spending on food, would be the dependent variable. The independent variables would be those that were thought to have some bearing on the amount a family spends on food, such as its price, household income and possibly some demographic variables relating to household composition (Whistler, 2008), (Claro and Monteiro, 2010). The parameters would then be estimated using statistical software on observed data. This single equation method can be used to predict household demand, study the cost and production functions of firms, as well as in labour and health economics (Studenmund, 2001). A variant of regression known as logistic regression will be used later in this thesis to obtain the transition probabilities for the dynamic microsimulation model.

One complication of using regression modelling is that economic variables are often interdependent so that the outcome of one process becomes an input to another, which in turn affects the first process. A classic example of this is the supply and demand cycle where supply affects demand and the resultant changes in demand go on to affect supply. In cases where there is a feedback loop of this kind, it is possible to use simultaneous equation methods estimated by Two-Stage Least Squares, to model the system (Studenmund, 2001). A comprehensive treatment of the principles, estimation and use of regression methods is provided by Intriligator (1978).

One of the difficulties with this approach is that the demand for any good is in principle, dependent on the prices of all other goods. As a result, when modelling a diverse household budget set, there would be far too many parameters to make

estimation possible. This can be alleviated somewhat by considering aggregate categories of goods such as clothing, housing etc. instead of the price for each individual item. However, this solution is not as straightforward as it might appear because the relationship between household consumption for an aggregate good and a set of aggregate prices is not necessarily the same as it would be when considering individual goods (Lutero, 2010).

3.4 Demand Systems

This approach to economic modelling uses a mathematical formula that links a set of independent or exogenous variables, describing the current conditions faced by the unit (individual, household, firm etc.), with a dependent or endogenous variable that represents the item of interest. This might be something like the number of hours worked in a labour supply model or the budget share of a commodity in an expenditure system. The number of parameters to estimate can be reduced by applying constraints derived from neoclassical economic theory. The general principles of which are summarised by (Weintraub, 2002) as:

1. People have rational preferences among outcomes;
2. Individuals maximise utility and firms maximise profits;
3. People act independently on the basis of full and relevant information.

In this view, economic agents attempt to gain the maximum satisfaction or utility from obtaining goods according to their preferences while taking into account any relevant constraints. They are presumed to continue obtaining goods until the cost of

acquiring another unit of the good exceeds the cost of obtaining it. Likewise, producers will sell more goods until the cost of producing them exceeds the amount consumers are prepared to pay. In this way, the supply of scarce resources is balanced by the demand from consumers.

3.4.1 Applications for Demand Systems

One of the most useful results from demand system estimation is the determination of the price and income elasticities of goods (Alpay and Koc, 2002). These show the rate of change in demand as price or income varies. With these in place, it is possible to predict the effect of a change in price or income. This is very useful for a commercial organisation in setting the price of goods or may play a role in determining the effects of a change in tax policy (Deaton, 1988). Alongside estimating demand elasticities, testing consumer theory has long been a main point of interest for demand systems (Boelaert, 2013). In addition to the theoretical motivation, a comparison of results with observed data can be used as an empirical test of the assumption of rationality. Another major area is to estimate the effect of economic changes on the consumer's utility or well-being (Cherchye et al., 2013). Welfare economics is a major area of study that has clear policy relevance as well as academic interest. Demand systems also can be used to investigate the effect of mergers between companies or the level of competition in a market (Lianos and Genakos, 2013).

3.4.2 Development of Demand Systems

A demand system can be constructed from the notion of a rational consumer who has

income y to spend on n types of product $q=(q_{1,}\cdots,q_n)$ that have prices $p=(p_{1,}\cdots,p_n)$. Income y and prices p are assumed to be exogenous or independent. The consumer chooses a combination of goods $q=(q_{1,}\cdots,q_n)$ attainable within income y such that $y=pq$ where utility $u(q)$ is maximised according to the customer's preferences.

It is quite possible to construct a demand system using a single equation for a single good. In this case, economic theory provides two restrictions that can be utilised to reduce the number of parameters to be estimated. One of these is the 'adding up' restriction which can be stated as: the sum of expenditure on all goods is equal to total expenditure. The other is known as 'homogeneity' which states: there is no change in demand if all prices and expenditure change proportionately.

A notable example of the single equation approach is Stone et al. (1954). This study of UK consumer's expenditure on non-durable goods was based on the demand equation:

$$\log q_i - \hat{e}\log(\frac{m}{p}) = \alpha_{ii}^{*}\log(\frac{p_i}{p}) + e_{ir}^{*}\log(\frac{p_r}{p}) + e_{is}^{*}\log(\frac{p_s}{p}) + \theta_i t + u_t$$

Where:

q_i is the budget share of good i

$\hat{e}$ is the elasticity of total expenditure

m is total expenditure and

p is a general index of prices so that

$\left(\frac{m}{p}\right)$ is an index of real income

α_{ii}^{*} is the compensated own price elasticity of good i

p_i is the price of the ith good

goods r and s are close substitutes or complements

θ_i is a parameter to be estimated

t is time

u_t is an error term

Further reductions in the number of parameters can be made by developing multi-equation systems because stronger, cross equation restrictions, can be applied. These are 'symmetry' where: the total substitution effect of a unit change in the price of one good relative to the other is the same as the effect of the other good on the former. There is also 'negativity' where the elasticity of goods is always less than 1. The first complete demand system was the Linear Expenditure System (LES: Stone, 1954).

This is derived from the Stone-Geary utility function, $U = \prod (q_i - \gamma_i)^{\beta_i}$

Where:

U is utility

q_i is the consumption of good i

γ and β are parameters

It can be shown that utility is maximised when:

$$q_i = \gamma_i + \frac{\beta_i}{p_i}(m = \sum_i p_i \gamma_i) \qquad i = 1, 2, 3, \ldots.n$$

and n is the number of goods.

The LES achieves a significant reduction in the number of parameters to estimate. In this case there are 2n - 1, due to the β_i and γ_i parameters and this is reduced by 1 due to the adding up constraint. This reduction is partly because the form of its utility function guarantees that all the general theoretical restrictions will be satisfied but also because it implies the additional restriction of 'additivity'. Sometimes known as 'want independence', additivity means that the marginal utility of any good is independent of the quantities consumed of all other goods. This restriction reduces the number of parameters to be estimated significantly, so allowing the study of a wider range of goods. However, this assumption becomes less plausible as the number of goods increases – the more goods there are, the greater the chance of interaction. Another consequence of additivity is that it is not possible to have 'inferior goods' where demand falls as total expenditure rises, despite the fact that this type of good has been observed in practice.

As the general restrictions of economic theory are implied in the derivation of the LES, it is not possible to test subsequently, whether they are supported by the data observed in any particular study. The Rotterdam Model (Barten, 1964) (Theil, 1965) by contrast, begins with a demand equation then tests whether the assumptions are valid. If they are, the restrictions can be applied to obtain a more precise estimate of the parameters.

Budget shares can be obtained in the Rotterdam model by:

$$w_i d \log q_i = \sum_j \pi_{ij} d \log p_i + \mu_i \sum_i w_i d \log q_i \qquad i = 1, 2, 3, \ldots n$$

Unrestricted estimation entails a large number of parameters, in this case proportional to the square of n due to the π_{ij} parameter. This limits the number of goods that can be considered. Also, it has been found that the model results in a rejection of the homogeneity and symmetry restrictions, along with the additivity restriction when it is applied (Barten, 1969). Similar results have been encountered using double logarithmic systems and this is thought to be due to treating the elasticities as constants (Thomas, 1987: 76). Such problems have led to the development of flexible functional forms such as the Translog model (Christensen, Jorgenson and Lau, 1975).

$$-\log U = \alpha_0 + \sum_i \alpha_i \log q_i + \frac{1}{2} \sum_i \sum_j \beta_{ij} \log q_i \log q_j$$

However, this still leads to an apparent rejection of some of the restrictions implied by the consumer theory (ibid).

One of the most widely used demand systems is known as the Almost Ideal Demand System (AIDS: Deaton and Muellbauer, 1980a) with its variants the Quadratic AIDS (QUAIDS: Banks, Blundell, Lewbel, 1997) and the semi-flexible Almost Ideal Demand System (Moschini, 1998).

The general model for the Almost Ideal Demand system is specified as:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\left(\frac{M}{P}\right) + \epsilon_i$$

Where:

w_i is the budget share of the ith good

M is the total consumption expenditure

p_j is the price of the jth good (j is a good other than i)

P is a price aggregator for the set of goods

ε_i is an error term for good i

It is possible to take the demographic characteristics of each household into account by including a vector of dummy variables Z, however this can greatly increase the number of parameters to be estimated. Each dummy variable indicates the presence or absence of the characteristic of interest. For example the characteristic of, 'age of the oldest household member', might be coded as 1: 'under 30' and 2: 'over 30'. This could be repeated for each characteristic of interest. The equation above would then take the form:

$$w_i = \alpha_i + \sum_j \gamma_{ij} \log p_j + \beta_i \log\left(\frac{M}{P}\right) + \delta_{iz} Z + \epsilon_i$$

The δiz indicates that the number of household parameters increases with the product of the number of goods and household attributes. As a result, it becomes more difficult to apply if the households are to be represented at a highly disaggregated level as they are in microsimulation modelling. This is the same issue that precluded the use of the cohort-component model for demographic microsimulation in Section 2.3.4.2. The problem is that as the number of dummy variables increases, the number of parameters becomes prohibitive due to data limitations. Also, the number of household categories is limited by the number of equations in the system so it would

not be possible to use continuous variables because this would imply an infinite number of categories. In many cases, the AIDS model also provides a rejection of consumer theory in that homogeneity and cross equation symmetry are rejected.

The Exact Affine Stone Index (EASI: Lewbel and Pendakur, 2009) is a demand system that is not restricted in its functional form and can also model the effect of unobserved heterogeneity (variation in spending behaviour due to factors not represented by the variables in the equation). Budget shares are given by:

$$w^j = m^j(y, Z) + \sum_{k=1}^{J} \alpha^{jk}(z) \ln p^k + \epsilon^j$$

where w^j is the budget share for good j

m^j is a set of functions of utility and observed household characteristics

y is implicit utility

z is a vector of demographic characteristics

p^k is prices for good k

ε^j represents unobserved heterogeneity for good j

Continuing work on demand systems has indicated that the inclusion of dynamic factors such as changing taste, habituation and consideration of the consumer's ability to anticipate future prices, may bring the models into agreement with economic theory.

3.4.3 Challenges for Demand System Analysis

The discussion of the development of demand systems shows that it is a very active area of research. This section highlights some particular problems and what is being done to solve them, that are of particular relevance to applying demand systems in the context of microsimulation modelling.

3.4.3.1 Dimensionality

One of the difficulties in demand system estimation is known as the problem of dimensionality. It was noted above, that in many types of demand system, the number of price parameters increases with the square of the number of goods and this makes it difficult to model disaggregated goods. One solution is to impose a multi-level budgeting scheme (Hausman, Leonard and Zona, 1994) (Hausman, 1996). Here the top layer could represent the budget shares of a market sector or aggregated good, while the lowest layer represents the differentiated goods of interest. Estimating each layer separately reduces the dimensionality of the problem. While this provides a way to make the modelling of differentiated goods feasible, it is done at the cost of making some subjective assumptions about product grouping and still does not increase the total number of goods that can be represented (Lianos and Genakos, 2013).

A more radical alternative to the dimensionality problem is to model the characteristics of goods, instead of the goods themselves (Lancaster, 1971). In this way, a few characteristics can combine to represent a large number of goods and closeness in characteristics corresponds to the likelihood of substitution among them.

In what are known as discrete choice models, the consumer selects one good from a bundle depending on the characteristics of the consumer and the goods (Nevo, 2001). The probability of selection can be calculated using a multinomial logit model (McFadden, 1973). This method is highly tractable but also contains the restrictive assumption that customers substitute in proportion to market share (Lianos and Genakos, 2013). A nested logit model can be used to alleviate this problem but here, the order of nesting makes a difference to the results; an issue which is addressed by Bresnahan, Stern and Trajtenberg, 1997). An alternative approach to modelling more flexible substitution patterns is the BLP framework (Berry, Levinsohn and Pakes, 1995). Their random coefficient multinomial logit model with unobserved product characteristics also includes observed and unobserved consumer heterogeneity and has been further developed by others such as Nevo (2001) and Petrin (2002). This provides a flexible framework for demand modelling. However, in some cases the characteristics of a product are subjective and difficult to define (Lianos and Genakos, 2013).

3.4.3.2 Functional Form

Many types of demand system use some kind of functional form to specify the consumer's demand behaviour. Whether it be linear, quadratic, Cobb-Douglas, Leontief or translog, economic theory gives little indication as to which is the most appropriate to use or which variables to include in the demand function (Goldberger, 1989). This raises the risk of misspecification (Blundell, Horowitz and Parey, 2012) and has led to the development of flexible functional forms which use a set of

parameters to allow for the effect of variations in response at different income and price levels. This alleviates but does not completely solve the problem of specification. One alternative is to use non-parametric methods based on the Theory of Revealed Preference (Samuelson, 1938, 1948) (Houthakker, 1950). This is founded on the Weak Axiom of Revealed Preference which states that if a customer chooses an option from a range of feasible choices, this indicates a stable preference for the chosen option. The Strong Axiom of Revealed preference adds transitivity to this definition and can be stated: if a consumer prefers A to B and B to C then they cannot prefer C to A. This provides a very useful tool for empirical nonparametric analysis of consumer choices (Varian, 2006). Non-parametric methods have received considerable attention by researchers and further developments have been made, such as The General Axiom of Revealed Preferences (GARP: Varian, 1982). Recent advances have considered extending the basic model to account for habit formation and multiperson households (Cherchye et al., 2013). In practice, nonparametric methods have sometimes given erratic and implausible results such as demand rising with prices, which is inconsistent with consumer theory. This has led to further research into imposing constraints on the results to restore compatibility with theory (Blundell, Horowitz and Parey, 2012). However, Moschini and Moro (1996) note that the data requirements of non-parametric methods make implementation difficult and may make the potential advantages difficult to achieve in practice.

3.4.3.3 Observed and Unobserved Heterogeneity

Econometric demand systems like the AIDS model described above usually consist of

two parts. The first, is a systematic component that is a function of observables such as price, income and demographics. The second is an error term which is independent of the observables but is different for each household. In many cases, the observables explain no more than half the variation in budget shares (Pendakur, 2009). The rest of the variation is due to other factors such as measurement error and unobserved heterogeneity, which is represented in the error term. Variation in the observables is reflected in changes to the budget shares but since the error term is a constant, the effect of the unobservables is not reflected in the marginal change in budget shares (Christensen, 2007). As a consequence, the model does not correctly allow for any unobserved heterogeneity because the thing that is being estimated in a demand system is the average demand conditional upon observed characteristics (ibid). The significance of this in affecting the results of demand modelling is noted by Lewbel and Pendakur (2013) for example.

Christensen (2007) shows how to address this problem by developing a demand system that allows for unobserved heterogeneity by including household specific intercept and slope parameters. This is done using a longitudinal panel survey of Spanish consumers. This framework demonstrates how the AIDS model can be extended to allow for unobservable heterogeneity but it is only possible due to the availability of the long panel survey. In most cases, only a single cross-sectional dataset is available which limits the general applicability of this innovation. The Exact Affine Stone Index (EASI: Lewbel and Pendakur, 2009) is able to account for unobserved heterogeneity and is flexible in its functional form. However,

dimensionality is still a problem due to the number of parameters to estimate. McAleer, Medeiros and Slottje (2008), for example, were restricted to six goods in their study of US food consumption.

3.4.3.4 The Rational Agent

It was noted above, that demand systems are usually based on the concept of the utility optimising rational agent. This allows for a high level of flexibility in modelling behaviour because it provides a way to anticipate how the economic unit adapts to its environment. It is possible to predict reactions to novel situations and this partly explains the diversity of application areas for demand systems that were mentioned above. Despite these advantages, the idea of the utility optimising rational agent has been questioned from a number of quarters. Kahneman and Thaler (2006) for example find the assumption implausible and identify four situations where people fail to maximise their experienced utility: 1) where the emotional or motivational state of the agent changes over time; 2) when the circumstances under which a decision is made are different from those pertaining when it is experienced; 3) when choices are based on flawed evaluations of past experiences and 4) when people must forecast their future adjustments to new circumstances.

Experimental tests of rationality in economic decision making have also revealed differences between what is observed and what would be predicted on the basis of rational behaviour. Gintis (2000) finds that real people have a higher propensity to cooperate than pure self interest would suggest. They also have been observed to

retaliate against free-loaders when there is no apparent gain or even at some personal cost.

Supporters of the neoclassical paradigm might respond by pointing out, as Mattei (2000) does, that the neoclassical model was never intended to represent the idiosyncrasies of an actual person. The neoclassical model of the human agent consists of a hypothetical representative consumer whose behaviour corresponds to that of a *group* of real consumers (Hicks, 1955). While the group can be defined to represent a very specific type of consumer, this is at odds with the essence of microsimulation modelling which is, by definition, concerned with representing individual cases. Others such as Friedman (1953) argue that a theory should be judged on the basis of its predictive power and not on the realism of its assumptions. It can also be noted that the assumptions of neoclassical economics are evolving in response to the critical attack. One notable example is Simon's (1957) ideas on 'bounded rationality' where, in the face of complex decisions and incomplete information, agents attempt to simplify their choices and seek a satisfactory solution which is not necessarily optimum.

3.5 Random Assignment

The previous section outlined some of the current problems associated with demand systems and what is being done to solve them. This thesis does not address them directly, rather it aims to circumvent them, for the purpose of microsimulation modelling, by implementing a method known as random assignment. The idea of

random assignment is usually associated with selecting individuals for treatment groups in such a way that the effect of the treatment is the only source of difference in outcomes between the groups. However, in the context of microsimulation modelling, random assignment is a kind of matching or imputation technique where a donor is selected on the basis of its similarity or closeness to the receiving unit.

This approach has been used by (Klevmarken et al., 1992) and (Klevmarken & Olovsson, 1996) to predict the type of house a family would buy in terms of size and cost, as part of a model of the Swedish housing sector called MICROHUS. Also, Hussenius and Selén (1994) used it to link together short panels of income data in a study of tax and transfer changes. More recently, Holm, Mäkilä, and Lundevaller (2009) used a random assignment scheme in a dynamic spatial microsimulation model of geographic mobility. They found that this approach had the potential to provide better population projections than the alternative interaction based models. However there has been no attempt to use a random assignment scheme as a way to model spending patterns.

In his survey of behavioural modelling in microsimulation, Klevmarken (1997) noted that the advantages of random assignment are that it is not necessary to impose a functional form on the data or make any assumptions about the distribution of variables. There are no parameters to estimate and the method preserves the variation and most of the correlation present in the original dataset. The approach also allows the study of situations where people behave in fundamentally different ways; in

particular where some individuals do something other than maximise their utility function.

Some of these features of random assignment are exploited in this thesis to develop an alternative to demand system estimation, for use in microsimulation modelling. The first of these is that random assignment does not require the estimation of parameters. Modelling takes place at the level of individual units so the heterogeneity represented in the model is the same as that is contained in the available data. Unobserved heterogeneity is embedded in the expenditure values themselves and so is implicitly taken account of in the copying procedure. As Andreassen (1993) noted, microsimulation models are a natural framework for incorporating relationships which take account of unobserved heterogeneity by using a probability distribution between observed variables that reflects the effect of the unobserved variables. In the models described later, the base dataset perform this role because its expenditure variables contain the result of unobserved heterogeneity in household characteristics and the copying procedure draws from this distribution. Second, there is no limit to the number of goods that can be modelled because the copying process can include as many variables as are available from the donor case. It is therefore not limited by the dimensionality problem. Third, there is no assumption of functional form. The relationship between income and expenditure (the Engel curves) is implicit in the data. This is also the case for the way other characteristics relate to each other such as age and expenditure. Fourth, the method is not tied to the assumption of the utility optimising rational agent. Behaviour is embedded in the expenditures and behavioural

change is implemented by copying from a similar case that has already experienced the new conditions. These characteristics suggest that this approach appears to have all the requirements for a method that can model a complex budget set at the micro-level without encountering problems due to dimensionality, functional form, heterogeneity or a predefined behavioural framework which may be at odds with the observed data.

Random assignment is related to statistical matching, which is used in microsimulation to combine data from a number of different sources or files. Here, individual cases are matched on the basis of the similarity between one or more variables which are common to both sources (Ingram et al., 2001). However, the random assignment scheme applied in this thesis is essentially different from statistical matching because records are matched from within the same dataset. This makes it more akin to hot-deck imputation where missing variables are obtained from a similar case that has a complete set of variables (Andridge and Little, 2010). Hot deck imputation is often used in surveys, where respondents may decline to answer some or all of the questions. This is particularly problematic because different respondents may have different propensities for omitting data so leading to biased results. A range of methods has been developed to ameliorate this problem such as propensity score matching (Rosenbaum and Rubin, 1983) which allows for the bias by estimating the probability of data being missing. Random assignment takes a different approach and is analogous to performing a hot-deck imputation where all of the data for the following year is missing. However the essential difference is that in

random assignment, the rate of missing data in any particular year is not changed. If there is any missing data in the donor case, the corresponding variables will still be missing in the recipient after copying. This means that throughout the dataset, the amount of missing data will not change unless, by chance, cases with missing data happen to be overwritten or additional cases with missing data are copied. For this reason, it seems appropriate to distinguish this approach from an application of hot-deck imputation and this thesis uses the name given to it by Klevmarken and others mentioned above who have applied this technique in different areas.

3.6 Advantages and Disadvantages of Demand Systems and Random Assignment in Microsimulation Modelling

The following table uses the evidence collected in this review of demand systems and random assignment to evaluate the potential strengths and weaknesses of these two approaches in the particular context of the microsimulation of household expenditure where the emphasis is on modelling cases at the level of individual units.

Demand Systems		Random Assignment	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
Flexibility in behaviour: the rational agent can make decisions to adapt to new situations.	Limit on disaggregation of economic units makes it difficult to take account of the effect of unobserved heterogeneity over time.	Represents economic units at the micro-level so accounts for both observed and unobserved heterogeneity of cases as they change.	Inflexibility in modelling behaviour: The copying procedure limits behaviour to previously observed cases.
Based on theory: economic theory provides framework for analysis.	Limit to the number of goods modelled (dimensionality).	No limit on the number of goods modelled.	No general approach to validation.
Validation: range of statistical validation tests have been used.	Need to specify functional form (except in unconstrained nonparametric approaches).	Imputation method familiar to microsimulation modellers.	Untested in expenditure modelling.
Errors can be estimated using a range of methods.	Tied to rational agent.	Ongoing research into matching and imputation as well as microsimulation.	
Ongoing research: methods are tested and built upon.	Requires specialised economic knowledge not usually possessed by microsimulation modellers.	Flexible output without respecifying model.	
		Can model interaction between units	

Table 2: Advantages and Disadvantages of Modelling Approaches

Table 2 shows that demand system analysis has several important advantages. It is a mature scientific paradigm based on a well defined theory and a set of methods and standards which have been developed incrementally over several decades. In consequence, it is easy to see why microsimulation has incorporated the econometric paradigm so widely. Meanwhile, random assignment has some serious drawbacks; a lack of flexibility in modelling behaviour and difficulty of validation. Nevertheless, this thesis is concerned with the specific area of the microsimulation of household

expenditure so that the economic component will be run in conjunction with a microsimulation. It was found in Chapter 2 that microsimulation already has the problems of a lack of flexibility in modelling behaviour (compared to ABM) and that it is also difficult to validate. In this context, adding a random assignment component to a microsimulation will add no new problems to those that already exist. In contrast, adding a demand system to the microsimulation would introduce the disadvantages of demand systems to the microsimulation. The principal one of these is to limit the level of disaggregation that can be represented as the model runs dynamically over time. This runs counter to the reason for operating a microsimulation model in the first place and is the same reason that the cohort component model was rejected as a way to model demographics in Chapter 2.

This thesis does not claim that random assignment is a superior method for economic modelling in general. However, within the context of a pre-existing microsimulation, it adds no new problems to those that already there and it avoids importing the difficulties which are, in this context, highly significant, which are associated with demand system analysis.

3.7 Conclusion

This chapter has described the development of parametric demand systems. Each new implementation was meant to improve upon current approaches and this process continues in an active research programme today. However, several difficulties were found to remain when modelling at the micro-level. One of the arguments of this

thesis is that random assignment can form the basis of an approach to expenditure analysis that does not limit the number of goods that can be considered and also does not loose information by aggregating households into groups. This will be tested in Chapter 5, which uses a random assignment scheme in an application to model the effect of changes in income on household expenditure patterns.

The end of this chapter marks the conclusion of the literature review phase of the thesis. Chapter 2 considered current methods for micro-level modelling and highlighted the problem of software complexity in microsimulation. The current chapter reviewed expenditure modelling and found that there are difficulties in representing the distribution of units at the micro-level. The next two chapters investigate the proposed solutions to these problems. Chapter 4 develops and evaluates a dynamic microsimulation model using NetLogo and Chapter 5 tests random assignment in the context of modelling the effect of income changes on household expenditure patterns. Chapter 6 will combine the two approaches in a substantive application to model the effects of population ageing on aggregate household expenditure and Chapter 7 investigates how random assignment can be applied more widely by describing how it was used in a commercial environment.

Chapter 4: NetLogo Demographic Model

4.1 Introduction

The previous two chapters provided a review of micro-level modelling and expenditure analysis with the aim of identifying gaps or limitations in current methods and tools. In the case of micro-level modelling, the problem was found to be in the complexity of developing the software for dynamic microsimulation and it was suggested that NetLogo could be used to make this process less burdensome. In expenditure analysis, it was argued that parametric demand systems are restricted in their ability to dynamically model large budget sets at the micro-level, due to data limitations in estimating the parameters for the equation. A random assignment scheme was proposed as a method that does not have this limitation and is also able to retain the information on the distribution of cases as they change over time.

This chapter moves beyond reviewing the literature and identifying problems, to applying the research methods described in Chapter 1 and obtaining results that begin to answer the questions that form the motivation for this thesis. In Chapter 1, it was stated that a practical approach would be taken, that involved developing models using NetLogo and using the results to determine its suitability as a platform for microsimulation modelling. This process begins here with the design, construction, validation and evaluation of a dynamic microsimulation model using NetLogo, known as Tyche. The chapter provides a detailed description of the modular design of Tyche and how the transition probabilities were obtained from survey data, drawing

on existing models where appropriate.

In addition to its objective of testing NetLogo as a platform for dynamic microsimulation, the model will also be used in Chapter 6 to simulate the effect of demographic change on household expenditure patterns. It is necessary therefore that the projections produced by the model are a plausible representation of the likely trajectory of demographic change assuming that current trends continue. The next part of the chapter validates model projections using methods described by Morrison (2008). Following this, Tyche is compared against one of the leading current implementations of a dynamic microsimulation framework, LIAM2 on the criteria of functionality, processing speed, usability and flexibility. This contributes further, towards the information needed to answer the research questions stated in Chapter 1, the analysis of which will be left until the concluding chapter.

4.2 Design Overview

The dynamic microsimulation model developed in this thesis follows the DYNASIM template, that was introduced in Chapter 1, with modules for births, deaths partnership formation and dissolution. On top of this basic structure, since the ultimate aim is to model the population at the household level, modules were added to represent single people leaving and returning to the parental home. A schematic overview of the model is shown in Figure 5.

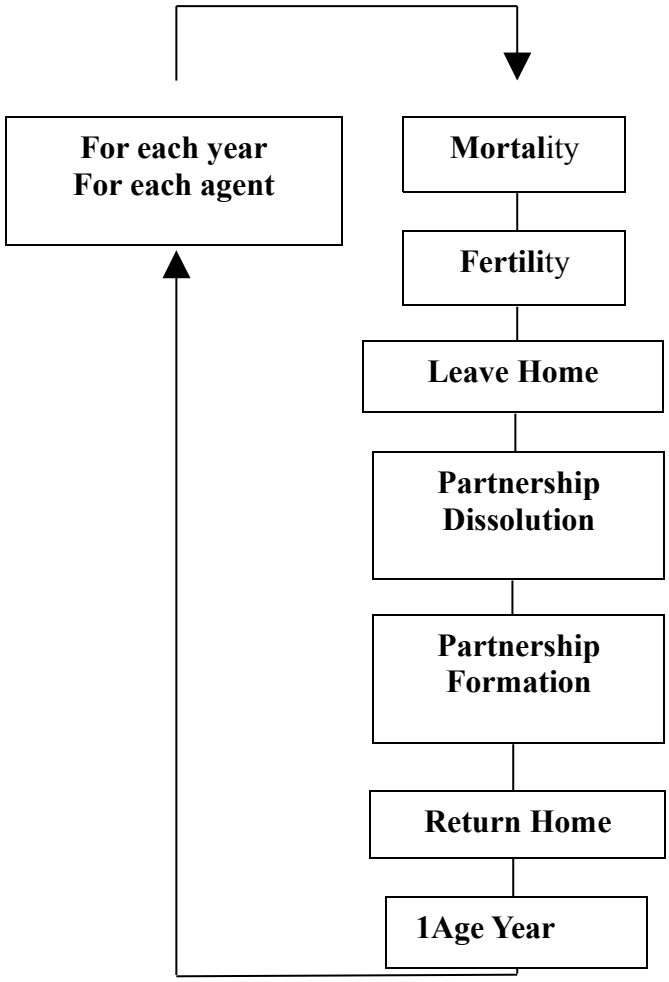


Figure 5: Microsimulation Modules

From figure 5, it can be seen that the microsimulation operates on an annual cycle and within that, steps through individuals in sequence. The latter will be done in a random order. The individual-level transitions, implemented in the sequence of modules, result in changes to household composition and so make it possible to project the evolution of the population at the household-level.

4.3 Base Survey Data

The initial cross-sectional population and the annual transition probabilities for the key events of births, partnership formation and dissolution were estimated using data from the British Household Panel Survey (BHPS). The BHPS is conducted by the ESRC Longitudinal Studies Centre and the Institute for Social and Economic Research (ISER) at the University of Essex. Its objective is to further the understanding of social and economic change at the individual and household level in Britain. The survey is conducted annually and aims to obtain a representative sample of the UK population. The sample is large, covering over 5,000 households and 10,000 individuals. Data is collected primarily by a questionnaire administered by face to face interview. At the level of the household, the survey includes a list of household members and their date of birth, sex, marital and employment status. A detailed individual questionnaire is administered to every adult member of the household. This includes sections on health and caring, demographics, earnings, employment as well as values and opinions. If panel members cannot be located in one particular year, a proxy schedule may be administered so that another person can record details of health, employment and income. The first wave of the survey was collected between September 1991 and April 1992. The bulk of the data collection takes place before Christmas. The sampling procedure is a two stage stratified systematic sample. A large volume of data is collected each year and held in a caseless SIR database which consists of several record types. Examples of these are HHRESP – household level information, INDRESP – individual level responses, JOBHIST – employment history, HHSAMP – interview outcome and weighting,

INDALL – data from household composition questions and INCOME – income and payment data. For each successive year of the survey, a prefix letter is added so the individual responses for the first wave would be held in the record type AINDRESP, the second, BINDRESP and so on. A lower-case 'w' is prefixed to the record type or filename to indicate the file without specifying a year e.g. wINDRESP.

The wide range of data within the BHPS made it possible to obtain all the variables needed to create the base dataset from one source and therefore circumvent the need for imputation from other sources. Imputation can introduce errors into the model and avoiding it ensures that this particular type of error is not present here. The process used to create the base dataset was that first, the chosen variables were selected from the BHPS using SPSS menus and stored in a new SPSS file. Missing values were recoded as appropriate, for example setting codes for missing data from -9 to zero. This file could then be saved as a plain text file compatible with NetLogo's file input primitives.

4.4 Obtaining Transition Probabilities

The annual transition probabilities are implemented in a set of logistic regression equations. In this type of regression, the variable to be predicted is dichotomous. In this case whether the transition occurred during the current year or not. The model also includes a set of independent variables which are thought to have some influence on the probability of the transition occurring. These are linked by the formula:

$$\text{logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots B_k X_k \quad (1)$$

where p is the probability of the transition, logit(p) is the logarithm of the odds of the transition. X_1 to X_k are the independent variables and b_0 to b_k are coefficients. These can be estimated using standard statistical software. When the coefficients have been determined, the numerical probability for the transition, given current values of the independent variables is:

$$p = 1 / 1 + e^{-\text{logit}(p)} \qquad (2)$$

Each transition takes place with a probability that depends on a number of factors. The determination of what these factors are is not a trivial exercise. For example, the selection of predictor variables for the probability of marriage in the coming year might begin with a review of previous work on the subject. This could be augmented by some statistical tests to find the most significant factors and whether they are likely to be stable over time. Finally, the regression equation could be estimated using a suitable longitudinal dataset. This represents quite a significant project in itself and a similar process would have to be repeated for every logistic regression equation in the model. Fortunately, the developers of previous microsimulation models have already been through this process and some have made their findings available to other researchers. Zaidi and Rake (2001) provide a useful summary of the design choices made in some microsimulation models. In MOSART for example, mortality depends on age, sex, marital status, educational attainment and disability. In DYNASIM2, marriage depends on factors such as age, race, sex and previous marital status. The design choices underpinning SAGE are described in detail in the SAGE Technical Notes (Scott, 2003) and were a key resource for the NetLogo based model.

Some of the annual transitions are relatively rare. For example, the number of males between the ages of 20 and 30 in the British Household Panel Survey (BHPS) who marry each year is typically in single figures. To obtain a sufficient number of these events for the derivation of an accurate annual transition probability, data from the first 15 years were pooled and weighted as needed using the provided BHPS longitudinal weight (wLRWGHT) to allow for differential attrition throughout the course of the survey (Taylor et al., 2009). The process used to estimate the coefficients for the regression equations was first to create a data file that contains information on each individual for pairs of years over the 15 year sample. Each row in this table holds items such as marital status in the first year and what it was in the following year; age of the youngest child in the first year and age of the youngest child in the following year. From this, it is possible to tell whether a transition occurred during each particular year and so estimate the probability of a transition occurring given the individual's state in the first year. There can be up to 14 pairs of years for each person, depending on how many of the years of the BHPS they participated in. The next step was to run the binomial logistic regression to obtain the coefficients for use in the microsimulation. This was done using an SPSS syntax file.

Subsequently, as each agent is processed during a simulation, its state variables, along with the coefficients, can be used to determine a numerical value for the logit as in equation 1 above. Next, the actual probability can be calculated by substituting this value into equation 2. The probability is then used to determine whether the transition occurs on this occasion by comparing it to a number drawn at random from a uniform

distribution between 0 and 1. If the random number is less than or equal to the probability of the event, the transition occurs. If it is greater than the calculated probability, nothing happens.

This probabilistic selection means that the fate of each particular individual depends on chance so that if the model is run again, with the same data but with a different seed in the random number generator, the same person will undergo a different trajectory – marry at a different time, have more or less children and have a different life span. The microsimulation gains consistency at the population-level because of the large number of individuals who are subject to the calculated transition probabilities. In the models described in this thesis, there are several thousand individuals so that within each module, the random draw is repeated several thousand times in each simulated year. The average number of transitions per simulated year is determined by the transition probability. The more likely the event, the more of those events that will happen each year. However, since the population is finite, there will still be some stochastic variation which decreases with the size of the population. In a population of several thousand, due to the central limit theorem, the distribution of the variation will closely approximate a normal distribution (Andreassen, 1993). This property of microsimulation is used later on to obtain confidence intervals for the distribution of outputs obtained over several simulations.

4.5 Operation of the model

4.5.1 Loading the Base Data

The initial population is stored in two files. One of these contains data at the household level. The other has data about individuals. The household file is read first. This sets up households identified by a unique case_number ready to be occupied by the agents. Next the individual (agent) file is read. Each agent moves to the appropriate household indicated by the respective case_number. Children are assigned to the parent they will be dependent upon in the event of the couple separating. In 9 out of 10 cases, following the SAGE model, this is the female. In the birth module, for simplicity, all offspring depend on their mother.

The BHPS provides a cross-sectional household weight (wHHWGHT in wave A and wXHWGHT for subsequent waves) to adjust for non-response and unequal selection probabilities at the household level. These are used so that the model can be interpreted as a representation of the UK population. A difficulty arises here because the weights are decimals but it is not possible to have a fraction of a household. This was dealt with by creating or deleting an entire household with a probability that depends on the weight. For example, if the weight for a particular household was 0.5, there would be a 50% chance that the household would be deleted; otherwise, it would be unchanged. If the household weight was 1.1, there would be a 10% chance that the household and its occupants would be duplicated. In this way, the now partly synthetic population was balanced to correspond with the real UK population.

4.5.2 Mortality

The BHPS provides a number of variables, such as xLVWHY in wINDSAMP that indicate the destination of former participants who have left the survey. One of the reasons for leaving is that the respondent is deceased, however, in some cases, it is not possible for BHPS staff, when they arrive to conduct the survey, to verify that the respondent has died during the previous year and so they are recorded as non-contacts (Uhrig, 2008). This would introduce a small negative bias in mortality rates derived using BHPS data. Also, those who are recorded as having died during the previous year do not have longitudinal weights for the current year (Sefton, 2007). This is a problem because it is these weights that are used here to calculate the transition probabilities. Fortunately, the Office for National Statistics (ONS) provides the annual probability of death for males and females which are derived from death certificates and this was considered to be a more reliable source of data, for this particular variable, than estimating it from the BHPS. The figures used for the microsimulation were taken from the ONS Life Tables 2004 – 2006 (ONS, 2013a). The effect of this is to freeze mortality rates at these values, which will introduce errors into the model if observed rates change from what they were originally. The model allows these probabilities to be changed and in the applications described later, mortality rates are amended or aligned to allow for ONS projected variations. Figure 6 shows the probability of an individual dying in a particular year against their age.

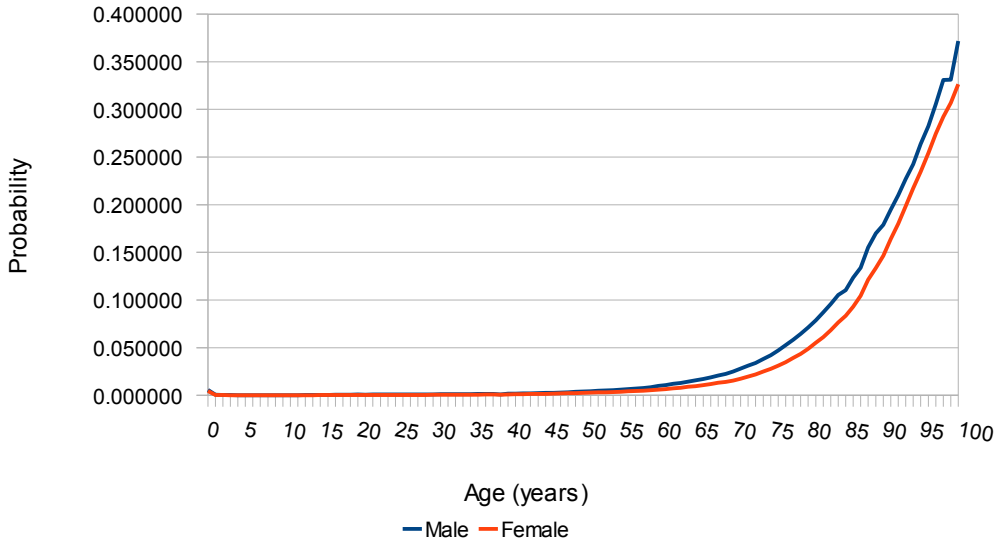


Figure 6: Risk of Death by Age (ONS)

It is apparent that the risk of death is relatively low until the age of 40. After this time, it rises slowly at first, then increases rapidly until by the age of 100, there is about a 30% chance of death during the year for women and a slightly higher chance for men. There are no data beyond the age of 100.

Due to the differences in the risk of death during the life course, the raw data was partitioned into two sections. For those aged under 40, the annual probability of mortality **p**, was approximated by taking the mean average over the time in question. This was obtained using Microsoft Excel and it was found that for males **p = 0.000736** and for females **p = 0.000411**. For those over 40, the annual probability of mortality was fitted with an exponential function again using Microsoft Excel and for males **p = 0.00003 exp(0.0954 age)** and for females **p = 0.00001 exp(0.1017 age)**. As the exponential function increases without bounds, the maximum annual probability

of mortality was capped at 0.5 to prevent it exceeding 1. (In a later version of this model, the actual values from the ONS Life Tables were used with no discernible difference to the results).

4.5.3 Birth

This module is applied only to females aged between 15 and 60. A review of previous microsimulation models (Chapter 2) indicated that the probability of a woman giving birth in the following year depends on many factors including her age, current number of children and the age of the youngest child. Current marital status is also significant and three separate logistic regression equations were estimated; one for married women, one for cohabitors and one for unpartnered women. Parameters for the models were derived using data from a pooled sample of the BHPS. Pairs of years from waves A (1991) to wave O (2005) were combined to provide an adequate sample size. The following tables show the parameter estimates for the three logistic regression models.

	β	S.E.	Sig.
Age	-0.155	0.005	0.000
Number of previous children	-0.264	0.460	0.000
Age of the youngest child	-0.020	0.008	0.829
Constant	2.875	0.177	0.000

Table 3: Parameter estimates of whether a married woman gives birth in the current year

	β	S.E.	Sig.
Age	-0.078	0.010	0.000
Number of previous children	0.168	0.085	0.048
Age of the youngest child	0.010	0.017	0.543
Constant	-0.440	0.298	0.140

Table 4: Parameter estimates of whether a cohabiting woman gives birth in the current year

	β	S.E.	Sig.
Age	-0.031	0.060	0.000
Number of previous children	0.466	0.054	0.000
Age of the youngest child	-0.021	0.009	0.019
Constant	-3.192	0.170	0.000

Table 5: Parameter estimates of whether an unpartnered woman gives birth in the current year

When an agent is selected to give birth, a function is called which creates the offspring and initialises its state variables. This is done using NetLogo's HATCH primitive. HATCH creates a specified number of agents, (one in this case) with state variables that are identical to those of the calling agent and it appears at the same location on the world grid. Following this, a set of commands is executed to modify some of the new agent's state variables. These include setting its age to zero and its marital status to single. The parent's WHO number (unique agent identifier) is recorded as well as the coordinates of its current location. The new agent's sex is also set at this point. Using the values given in the SAGE documentation, there is a 48% chance of the agent being female and a 52% chance of it being male.

Probabilities of twins or triplets are modelled using data from SAGE Technical Note no. 4, which was in turn derived from (Office for National Statistics, 1996).

Mother's age					
	<=19	**20-24**	**25-29**	**30-34**	**35+**
Twins	0.00626	0.00861	0.01177	0.01510	0.01767
Triplets or more	0.00000	0.00009	0.00024	0.00068	0.00085

Table 6: Probability of multiple births by age of mother

Following the creation of the first agent, the same random number is compared with the probability of having twins appropriate to the age of the parent agent. If it is less than or equal to this threshold, one more offspring is created by the process described above. After this, another comparison is made against the threshold for having triplets. If the number is less than or equal to this, a third offspring is created. Using SAGE data has the effect of fixing the probability of multiple births at 1996 levels, however as can be seen from the stated likelihood of these events in Table 6, a marginal change in these values is unlikely to have a significant effect on the overall results.

4.5.4 Leaving the Parental Home

This module is designed to represent cases where individuals leave the parental home without marrying or cohabiting. As was the case in the mortality module, leaving the parental home results in the respondent moving out of the scope of the BHPS. In consequence, there is no longitudinal weight for the current year, which would be needed for the estimation of the transition probability. Also, in many cases, the reason for the move is given as 'other' so that it is not clear whether the individuals in question are leaving to set up their own home or whether the move is temporary (they

will return in the current year) or permanent. These uncertainties could lead to an inaccurate estimate of the rate of new household formation. Consideration of this problem led to the conclusion that a proper study of this transition was not feasible in the context of the current project and that there was no alternative than to adapt the results of previous research for use here. The most appropriate available was Aassve et al. (2005) and the transition probabilities for leaving home were based on these values.

In the current implementation, to be eligible for this transition, individuals must be aged 18 or over, have a marital status of single, have no dependants and currently live in a household containing more than one occupant. For this group, there is a probability of 0.1 of being selected to leave the current household to set up a new household where they are the sole occupant.

4.5.5 Marriage dissolution

Separate logistic models for males and females determine which marriages end in the current year. Most dynamic microsimulation models are female dominated in that the transition probability is calculated for females only and male transitions occur as a consequence. However, this arrangement loses some information about the males as can be demonstrated in a hypothetical example. Suppose two men get married. One has been married and separated a number of times before. For the other, this is his first marriage. It may be that first male's separations indicate a propensity for this type of event and the probability of separation is greater for him than for the male

being married for the first time. A female oriented dissolution module would not take this or any other differences between the males into account. The current model addresses this by alternating at random between being female dominated and male dominated so that, at some point, the male's characteristics have some bearing on the outcome. Of course, when the module is male oriented, information about the female is lost, however, the proposed arrangement will avoid any systematic bias that may be introduced by basing the module entirely on one sex. A better solution would be to take the characteristics of both partners into account simultaneously but this would be more complex to estimate and the potential gains in accuracy may not be justified in the context of the current application.

The model that operates for the current individual is randomly selected to be male or female dominated with a probability of 0.5. If the current agent in a particular instance happens to be male and the model for this agent is selected to be female dominated, the male avoids being considered for marital dissolution. This effectively halves his probability of dissolution. This is compensated for when his partner is processed through this module as he is subject to the full risk of her dissolution probability when the model is selected to be female dominated.

The parameter estimates for marriage dissolution are shown below. For males, this transition is based entirely on discrete age bands with the reference level, indicated by an asterisk, being the youngest age group. For females, since the birth module has already been run, the presence of a new birth during the year in question can also be

taken into account.

	β	S.E.	Sig.
Age			
16-29*	0.000	-	-
30-39	-0.627	0.160	0.000
40-49	-0.954	0.165	0.000
50-59	-1.848	0.211	0.000
60-69	-3.078	0.353	0.000
Constant	-3.112	0.130	0.000

Table 7: Parameter estimates of marriage dissolution for males

	β	S.E.	Sig.
Age			
16-29*	0.000	-	-
30-39	-0.503	0.125	0.000
40-49	-1.036	0.139	0.000
50-59	-1.938	0.189	0.000
60-69	-3.378	0.376	0.000
Birth in the current year	-1.134	0.298	0.000
Constant	-2.947	0.099	0.000

Table 8: Parameter estimates of marriage dissolution for females

When an agent is selected for separation, its status is set to ‘separated’, and a ‘newlyseparated’ flag is set to prevent these individuals from re-marrying in the current year. It is reset at the start of the following year’s cycle. Next, the status of their partner is set to ‘separated’. With a probability of 0.9, the male moves to a vacant household space. In the remaining 10% of cases, the female moves out. In all cases, any children present, for the moment, remain at the current location and wait to be moved at the end of the current year.

4.5.6 Cohabitation dissolution

This module operates in the same format as the marriage dissolution module although the actual probabilities are calculated using separate logistic regression equations.

	β	S.E.	Sig.
Age			
16-29*	0.000	-	-
30-39	-0.671	0.146	0.000
40-49	-1.438	0.263	0.000
50-59	-0.611	0.253	0.016
60-69	-2.064	0.630	0.001
Constant	-1.932	0.081	0.000

Table 9: Parameter estimates of cohabitation dissolution for males

	β	S.E.	Sig.
Age			
16-29*	0.000	-	-
30-39	-0.523	0.134	0.000
40-49	-0.898	0.217	0.000
50-59	-1.115	0.292	0.000
60-69	-1.197	0.672	0.075
Birth in the current year	-0.526	0.258	0.042
Constant	-1.923	0.072	0.000

Table 10: Parameter estimates of cohabitation dissolution for females

4.5.7 Cohabitation formation

Separate probabilities are calculated for males and females with age and current marital status taken into account. In the female model, birth in the current year's fertility module is allowed for also.

	β	S.E.	Sig.
Age	-0.021	0.004	0.000
Current marital status			
Never married*	0.000	-	-
Widowed	-0.033	0.329	0.921
Divorced	0.670	0.133	0.000
Separated	1.210	0.153	0.000
Constant	-2.199	0.100	0.000

Table 11: Parameter estimates of a male beginning cohabiting in the next year

	β	S.E.	Sig.
Age	-0.045	0.004	0.000
Current marital status			
Never married*	0.000	0.000	0.000
Widowed	-0.690	0.279	0.013
Divorced	0.536	0.124	0.000
Separated	0.478	0.154	0.002
Birth in the current year	0.757	0.167	0.000
Constant	-1.357	0.104	0.000

Table 12: Parameter estimates of a female beginning cohabiting in the next year

The first step is to check whether there is a suitable partner of appropriate age available in the population. This is defined initially as individuals who are not already married or cohabiting, aged 16 or over and have an age that is within 20 years of that of the current agent. If a suitable agent cannot be found, the process is aborted and a warning message appears in the output window. In practice, this event has been found to be rare unless the population is very small. When the full BHPS population is used, this hardly ever occurs. If a suitable potential partner is present, a number is drawn at random from a uniform distribution between 0 and 1. If this is less than or equal to

the threshold given by the logistic regression equation, the process continues. If the number is greater than the threshold, no match is attempted. The next stage in the process is to search for a partner that has an appropriate age difference, selected using a probability derived from Bhrolchain (2005). This research indicated that in 2001, 9% of marriages were between a woman who was 5 or more years older than the man. In 6 % of marriages the woman was 3 to 4 years older than the man. In 11% of marriages the woman was 1 to 2 years older than the man and 9% of marriages were between individuals of the same age (or within 1 year). In 19% of marriages the man was 1 to 2 years older than the woman. In 15 % he was 3 to 4 years older, in 20% he was 5 to 9 years older and in 11% of marriages the man was 10 years or more older than the woman. As there seems to be no corresponding data available for the formation of periods of cohabitation, the same figures were used for marriage and cohabitation. This is not ideal because the age differences between members of the two modes of partnership formation may not be identical. Further research would be required to determine the precise effect of this assumption however it is not thought to be significant in the context of the models reported in this thesis.

Bhrolchain's figures allow a look-up table to be constructed to select the age of the partner for the agent that has been selected for cohabitation. For example, if the simulation was operating in female dominated mode, there would be an 11% chance of the partner being aged 10 or more years older than the female. In this case a random number between 0 and 1 would be drawn from a uniform distribution and compared with this probability. If it is less than or equal to the threshold, an

unpartnered male would be selected at random from the population who is 10 or more years older than the female (but not more than 20 years older). The next age band is a partnership where the man is 5 to 9 years older than the woman. There is a 20% probability of this. If the number is greater than the previous threshold of 0.11 and less than the cumulative total of the probabilities (0.31) then a male is selected who is 5 to 9 years older than the female. This 'roulette wheel' selection continues until one of the age bands has been selected.

If no suitable agent can be found within the desired age band, the procedure reverts to the default 20 years either way. This was found to occur in 18% of cases in the male dominated version and 6% of cases when it is the female that requires a partner. This will result in the age mixture of modelled partnerships being slightly more heterogeneous that is the case in practice. Finally, the partnership formation can be executed by the current agent moving to a vacant household and the partner moving to join them. Their marital status can then be updated to cohabiting.

4.5.8 Marriage formation

Couples who are already cohabiting are considered for marriage provided the cohabitation did not start in the current year. As the cohabitation module runs before the marriage module, new cohabitors must remain in this state until the current year's processing is complete to avoid depleting the proportion of cohabiting couples observed in the current year. This is mediated by the 'newcohabitation' flag. If selected, the resident couple set their status to 'married' and they stay at their current

location. Tables 13 and 14 show the parameter estimates for this transition.

	β	S.E.	Sig.
Age			
16-29*	0.000	-	-
30-39	0.015	0.117	0.896
40-49	-0.541	0.177	0.002
50-59	-1.001	0.279	0.000
60-69	-0.829	0.342	0.015
Constant	-1.807	0.078	0.000

Table 13: Parameter estimates of whether a cohabiting male will marry his cohabitee in the current year

	β	S.E.	Sig.
Age			
16-29*	0.000	-	-
30-39	-0.126	0.116	0.277
40-49	-0.121	0.160	0.449
50-59	-0.708	0.240	0.003
60-69	-1.113	0.630	0.077
Constant	-1.862	0.068	0.000

Table 14: Parameter estimates of whether a cohabiting female will marry her cohabitee in the current year

Individuals living separately are considered for marriage in a way similar to the cohabitation module. Probabilities for the age difference between members of the partnership are again taken from Bhrolchain (2005). In 11% of cases, in the male dominated version of this module and in 3% of cases in the female dominated version there was no partner of the preferred age available in the population. When this happens, a maximum age difference, set at plus or minus 15 years is applied. A separate set of logistic regression equations is used to select candidates for marriage

as shown in tables 15 and 16.

	β	S.E.	Sig.
Age	-0.001	0.006	0.834
Current marital status			
Never married*	0.000	-	-
Widowed	0.137	0.467	0.769
Divorced	0.692	0.222	0.002
Separated	1.684	0.218	0.000
Constant	-4.207	0.180	0.000

Table 15: Parameter estimates of whether a non-cohabiting male will marry in the current year

	β	S.E.	Sig.
Age	-0.028	0.006	0.000
Current marital status			
Never married*	0.000	-	-
Widowed	0.172	0.351	0.626
Divorced	0.686	0.209	0.001
Separated	2.100	0.176	0.000
Birth in the current year	0.644	0.304	0.034
Constant	-3.408	0.175	0.000

Table 16: Parameter estimates of whether a non-cohabiting female will marry in the current year

4.5.9 Age 1 Year

Age is incremented by 1 year ready for the next annual cycle. The current year is incremented.

4.5.10 Returning home after separation

Individuals whose cohabitation or marriage dissolved in the current year return to their parent's or original home if they are aged under 30 and have no dependants. It is assumed that those aged 30 and over will be able to find their own (unoccupied) accommodation. While this may not be true in practice, this group have a high probability of leaving in the current year and the overall effect is not significantly different.

4.5.11 Locating children

As agents change location due to partnership formation and dissolution, their offspring require a method of moving with them. This is mediated by a unique identification number that all agents have, called the 'WHO' number. Offspring store the WHO number of the parent it is designated as being dependent upon. This is recorded at the offspring's birth or in the 'findparents' module, which is run once after loading the initial data files. At the end of each year, all offspring check the location of this parent and move to the household that the parent currently occupies.

In some cases, about 20 each year, the parent cannot be located. This may be because the designated parent has died during the year or the offspring was not allocated a parent when the base dataset was loaded. When this happens, the offspring is adopted by a randomly selected 'family'.

4.5.12 Alignment

The issue of alignment was discussed in section 2.3.4.3 above. It was noted that, even well specified dynamic microsimulation models are generally quite poor at the unconditional forecasting of future macro-level parameters. This is due to the underlying transition probabilities drifting away from the values that were observed and faithfully recorded by the microsimulation model during its estimation. This is usually dealt with by adjusting transition probabilities so that model outputs are aligned with observed data, official projections or some hypothetical scenario. In the model described here, a simple alignment scheme can be implemented by altering the transition probability for births, deaths, marriage and cohabitation formation and dissolution, directly from the user interface or within the program code. This makes it possible to run alternative scenarios or reproduce values from official projections.

4.5.13 Summary of Transition Probability Determination

Table 17 provides a summary, for each transition type included the model, showing how it was calculated, who it applies to, which predictor variables are used and the source of the data for estimating it.

Transition Type	Applies to	Independent Variables	Source
Mortality (annual probabilities for males and females)	All	None	ONS Life Tables (2004 - 2006)
Birth (separate equations for married, cohabiting and unpartnered women)	Females aged 15 to 60	Age, number of previous children, age of the youngest child	BHPS 1991 - 2005
Second and third births	Females aged 15 to 60	Age, birth in the current year	SAGE
Leaving home	Aged 18 of over, single, no dependants	None	Aassve et al. (2005)
Marriage dissolution (separate equations for males and females)	All married people	Males: age Females: age and birth in the current year	BHPS (1991 to 2005)
Cohabitation dissolution (separate equations for males and females)	All cohabiting people	Males: age Females: age and birth in the current year	BHPS (1991 to 2005)
Cohabitation formation (separate equations for males and females)	Unpartnered people aged 16 and over	Males: age Females: age and birth in current year	BHPS (1991 to 2005) Age difference of partner (Bhrolchain, 2005)
Marriage formation (cohabiting) Separate equations for males and females	Aged 16 and over, cohabiting in the previous year	Males: age Females: age and birth in the current year	BHPS (1991 to 2005) Age difference of partner (Bhrolchain, 2005)
Marriage formation (unpartnered) Separate equations for males and females	People aged 16 and over: never married, widowed, divorced, separated	Males: age Females age and birth in the current year	BHPS (1991 to 2005) Age difference of partner (Bhrolchain, 2005)
Returning home	People aged under 30 who separated in the current year	None	Aassve et al., 2005

Table 17: Summary of Transition Probability Determination

4.6 Validation and Verification

In software engineering, is is usual to distinguish between validation and verification (Somerville, 2006). Validation is concerned with ensuring that the program performs its intended function (are we building the right product?). Verification is to do with checking that the program meets its specification (are we build the product right?). These two issues are sometimes conflated in dynamic microsimulation modelling so that verification is included within the validation process. Zaidi and Rake (2001) for example, suggest that validation incorporates both ensuring internal consistency and external credibility among the scientific community. For Caldwell and Morrison (2000), validation is a proactive diagnostic effort to ensure the models outputs are reasonable for their intended purpose and that the results are reasonable and credible. In this section, to fit with the microsimulation literature, validation will refer to the whole process of checking that the model is working correctly.

The purpose of validating Tyche is to ensure that it does what it was intended to do. This model has two main purposes. One is to assess the suitability of NetLogo as a microsimulation modelling platform. The exercise of developing the model provides some of the information for this and will be augmented later in this chapter when the model will be compared with LIAM2. The other is to produce disaggregated demographic projections for input into the expenditure module. These are not intended to be unconditional forecasts of population variables at some time in the future. As we have seen in the discussion on alignment (Section 2.3.4.3), microsimulation is not necessarily the best way to do this. However, the projections

are not arbitrary and the evolution of the population should represent a plausible trajectory, given the characteristics of the base population and their observed transition rates over the period during which the model coefficients were estimated.

4.6.1 Sources of error

From the conception of a model, which was defined earlier, its characteristics are that it is a simplification of the real system which abstracts those elements which are relevant for its purpose. The process of simplifying or abstracting from the real world means that there are several points where the model does not replicate its characteristics exactly. This can lead to errors where the model output does not correspond to what would be observed in practice (Cohen, 1991), (Klevmarken, 1998), (Wolf, 2001).

4.6.1.1 Measurement error

The BHPS is an exceptionally well designed survey, conducted by well trained professional interviewers. Nevertheless, it is impossible that every item of data is a fully accurate representation of reality. Respondents may be mistaken in their recollection or misunderstand a question. Also, the definition of variables must be defined or operationalised for collection. This could involve defining categories which may contain an element of subjectivity or grouping continuous variables into bands.

4.6.1.2 Sampling Variation

The demographic model uses the BHPS to obtain the base population and to estimate the annual transition probabilities. While the BHPS is one of the most comprehensive social surveys carried out worldwide, it represents a sample from the UK population and so is subject to some amount of sampling variation where the modelled population does not exactly match that of the UK. This will make a difference to the results because these depend on who is selected into the sample (Goedeme et al., 2013). Pudney and Sutherland (1994) show how to calculate the effect of sampling variation in the context of static microsimulation models but there is no generally applicable method that can be applied to more complex dynamic microsimulation formats (Goedeme et al., 2013).

4.6.1.3 Static Parameters

The transition probabilities were estimated from a pooled sample based on BHPS data. There is sampling variability here (albeit reduced due to the larger sample), which affects the values of estimated parameters but also the parameters for the equation are fixed at an average level for the estimation period. For most panel surveys, this rarely extends for more than a few years. If the probability of transitions change over time in the real population, model projections will drift away from observed values (Bacon and Pennec, 2009). This issue can be ameliorated by alignment, which is used here to compensate for changes in birth and death rates.

4.6.1.4 Limited Number of Parameters

Only a few independent variables were used in estimating the equations. While they were carefully selected and the parsimony of the model will reduce the incidence of multicollinearity, the model cannot reproduce the influence of the factors that were omitted. This will inevitably have an effect on the accuracy of model projections, which will be estimated later by comparing them with observed data. Also, several parameters were obtained from previously published sources, often obtained at different times. As a result, these probabilities will be fixed at a particular time and will not necessarily correspond with parameters estimated at different times. Due to the typically moderate and slow rate of change in demographic factors, these effects are assumed to be small compared to the sometimes precipitous swings in consumer demand which is the ultimate subject of this model. However, without updated sources for these parameters, this assertion cannot be tested.

4.6.1.5 Stochastic Variation

The next source of error or uncertainty derives from the stochastic nature of the simulation. A random draw is used to determine whether an individual transitions from one state to the next. This is repeated for all individuals and a different number of transition events occur in each simulated year. Therefore, since the random number generator is reseeded between runs, the simulation will produce a different result every time so that it is necessary to complete several runs to estimate the level of stochastic variation present. This also applies when the number of households is expanded after loading due to the weighting procedure. As the household is

duplicated according to a probability derived from its weight, it can be weighted up differently in each instance of a simulation. Also, household-level weights are used, which only take into account certain characteristics of the household (Taylor, et al., 2009).

4.6.1.6 Programmer Errors

Another potential source of error is due to program bugs. However thoroughly a piece of software is tested, it cannot guarantee the absence of defects because as (Dijkstra, et al., 1972) noted, the testing process can only demonstrate the presence of errors not their absence. Also, the number of outputs is so large in a disaggregated model that it is not possible to test every one.

4.6.1.7 Demographic Specification

An additional source of error is in the specification of the demographic processes. A range of simplifying assumptions have been made in restricting the range of demographic processes to the ones that are standard in the dynamic microsimulation literature. Incorporating other processes such as the formation of multiple family households, same sex couples and multi-generational households would add more sophistication to the model and probably make the results more accurate. However, the complexity of the demographic system to be simulated is virtually infinite and the resources available to complete the work are always finite. At some point, a decision must be made concerning what to include and what to leave out.

Another implicit assumption made here, is that changes in household composition are entirely dependent on the individual-level processes of births, deaths, leaving home, partnership formation and dissolution. It is quite possible for household-level characteristics to influence individual-level behaviour. For example, an overcrowded household may encourage individuals to leave when they would not otherwise have done so. This downward causation would form an interesting area for further research but is not implemented in this model in the interest of making it easier to understand and avoiding the emergence of feedback effects which could obscure the relationships between household composition and spending that is the aim of this research.

4.6.1.8 Imputation Error

In many cases, it is necessary to construct the base dataset by piecing together items from different sources. Here, errors can be introduced if the same variable is measured differently in the different sources. Also, matching records from more than one source is possible only if there are a number of variables which are common to both files. There may be some correlation between unmatched variables leading to bias in the matched file. Since all data was taken from the BHPS, this type of error is not present in the models described in this thesis but it would appear if a merged dataset was used to create the initial population.

4.6.2 Validation Scheme

In Chapter 2, it was noted that, compared to the work of generating results, only limited effort has been placed on validation (Li and O'Donoghue, 2013). Harding,

Keegan and Kelly (2010) find only one example (Morrison, 2008), of guidance on how best to validate complex dynamic microsimulation models. This will be used in the validation of Tyche by checking that the output is correct - as Caldwell and Morrison (2000: 202) put it, 'to assess whether the outputs are reasonable for their intended purpose'. Its application begins during the construction of the model, with the aim of minimising the introduction of errors, then it is used to measure the accuracy of model projections and evaluate whether they are acceptable for its intended purpose. Finally, where the errors are not acceptable, a process of alignment will be used to constrain key variables to appropriate limits.

Morrison (2008: 13) provides a framework for validating microsimulation models. It consists of several elements which are:

Data / Coefficient / Parameter Validation;

Programmer's / Algorithmic Validation;

Module-specific Validation;

Multi-module Validation and

Impact Validation.

Harding et al. (2010) describe how this framework was applied in the validation of the Australian Population and Policy Simulation Model (APPSIM: Keegan and Kelly, 2009). The validation process described below uses this case as a template. For each stage, the way it was applied for APPSIM is described first. This is followed by a description of how it was conducted for Tyche.

4.6.2.1 Data / Coefficient / Parameter Validation

The base data for APPSIM consisted of a 1% census sample from the Australian population; a dataset which was considered to be reliable and robust. The structure of modules was developed following a review of other dynamic microsimulation models. This helped to direct efforts to derive accurate and appropriate coefficients and parameters for the model. Transition probabilities were derived from seven years of the Household, Income and Labour Dynamics in Australia (HILDA) panel survey. This is a longitudinal survey which follows almost 20,000 individuals within over 7,500 households.

The base dataset for Tyche came from the BHPS which is a high quality longitudinal survey, weighted to form an accurate representation of the UK population. Fifteen waves were available for deriving transition probabilities. The design of modules and selection of parameters was informed by a review of existing dynamic microsimulation models, especially SAGE (Scott, 2003). Model coefficients were obtained using standard statistical techniques such as logistic regression or where available, from published sources such as the ONS Life Tables.

4.6.2.2 Programmer's / Algorithmic Validation

In APPSIM, the C# code was implemented by a specialist programmer while the modules were designed by members of the microsimulation team. This stage of the validation consisted of the designer reviewing the code written by the programmer to check that the original specifications had been implemented as intended.

Tyche was designed and coded by the author, who has postgraduate qualifications in computer science and has previously worked as a programmer. Validation was an ongoing process of ensuring that the code operated correctly, in that the values of relevant variables were as they should be. The high-level nature of NetLogo, combined with its model specific commands provided significant assistance in this process. The 'world grid', which gives a visual representation of individuals within households, along with the ability of 'probes' to view agent variables as the program runs, were of great help in checking that the program was working as intended.

4.6.2.3 Module-specific Validation

APPSIM has a comprehensive set of alignment facilities which enable it to reproduce external benchmarks where required. Individual modules were tested by setting alignment on, for all modules except for the one under test. In this configuration, any deviation from benchmarks would probably be due to the unaligned module and the source of any errors could be traced to that area.

Tyche was developed in sequence one module at a time. The initial modules were coded and tested to ensure they were operating as intended. As further modules and parts of modules were added, any new errors were likely to be due to the new code and so could be isolated and corrected relatively easily.

4.6.2.4 Multi-module Validation

The APPSIM team conducted an extensive set of comparisons between model output

and external sources such as Treasury projections and data from HILDA. APPSIM provides a range of tools to assist in this process. These include the 'Individual Output Tool' (IOT) which enables the reporting of every parameter associated with a specific individual. By looping through a set of individuals, output can be generated for the whole population, 'everyone output', or can be done for specific groups of interest. This can be extended over time to perform 'cohort tracking'. APPSIM also provides a wide range of summary statistics which can be sent to a spreadsheet and viewed in a set of graphs and tables.

In Tyche, multi-module validation took place at a number of levels. These included an inspection of individual output to ensure that key demographic variables were updated correctly. There was also some cross-sectional and longitudinal output checking against BHPS and ONS data, both with and without alignment. The results from each of these stages is described in the next section.

4.6.2.5 Individual Output

NetLogo provides file output commands that make it easy to save the details of each case for later inspection. The output is voluminous for a large population so only a few example cases are reproduced below.

The following table was built up from results collected at 5 year intervals during a simulation. They show details of the oldest occupant of the household, which corresponds approximately to the Household Respondent Person (HRP) as used in the

BHPS. This is done for a selection of households and also shows the relationship between household members. Four types of household composition were defined: a household containing one person is denoted as ‘single person’; a household with two occupants where the oldest person is either married or cohabiting is designated as a ‘couple’. If there is one member of the household who is aged 16 and over and one or more members aged under 16, this is considered as a ‘single parent’ household. All remaining households are classed as ‘other’.

2006

Household ID	No. Persons	Sex of Oldest Person	Age of Oldest Occupant	Composition	0-4.	5-17.	18-44.	45-64.	65
1004	3	male	23	other	1	0	2	0	0
1011	4	female	33	single parent	0	3	1	0	0
1016	4	male	42	other	0	1	3	0	0
1017	4	male	47	other	0	2	0	2	0
1047	3	male	55	other	0	0	1	2	0
1048	2	male	76	couple	0	0	0	0	2
1049	3	male	43	other	0	1	2	0	0

2011

Household ID	No. Persons	Sex of Oldest Person	Age of Oldest Occupant	Composition	0-4.	5-17.	18-44.	45-64.	65
1004	5	male	28	other	2	1	2	0	0
1011	4	female	38	other	0	2	2	0	0
1016	4	male	47	other	0	0	3	1	0
1017	3	male	52	other	0	0	1	2	0
1047	2	male	60	couple	0	0	0	2	0
1048	2	male	81	couple	0	0	0	0	2
1049	4	male	48	other	1	1	1	1	0

2016

Household ID	No. Persons	Sex of Oldest Person	Age of Oldest Occupant	Composition	0-4.	5-17.	18-44.	45-64.	65
1004	5	male	33	other	0	3	2	0	0
1011	3	female	43	other	0	0	3	0	0
1016	3	male	52	other	0	0	1	2	0
1017	2	male	57	couple	0	0	0	2	0
1047	2	male	65	couple	0	0	0	1	1
1048	1	female	85	single person	0	0	0	0	1
1049	4	male	53	other	0	1	1	2	0

Table 18: Individual Output

From Table 18 it can be seen that household 1004 has 3 occupants in the base year of 2006. The oldest is a 23 year-old male. The household also has two occupants aged between 18 and 44 and one in the 0 to 4 age band. Five years later, the male is 28 and there are now 5 occupants. The two new occupants are aged 0 to 4 while the one

previously aged 0 to 4 is now in the 5 to 17 age band.

This kind of table provides a rich source of data and inspection of individual output facilitates checking that the individual and household variables are updated as intended and that household composition changes in a coherent way over time.

4.6.2.6 BHPS Cross-sectional Validation

Figure 7 shows an example which compares the projected household size distribution in 2008 with that observed in the BHPS. Discussion of the results is left until Section 4.7.

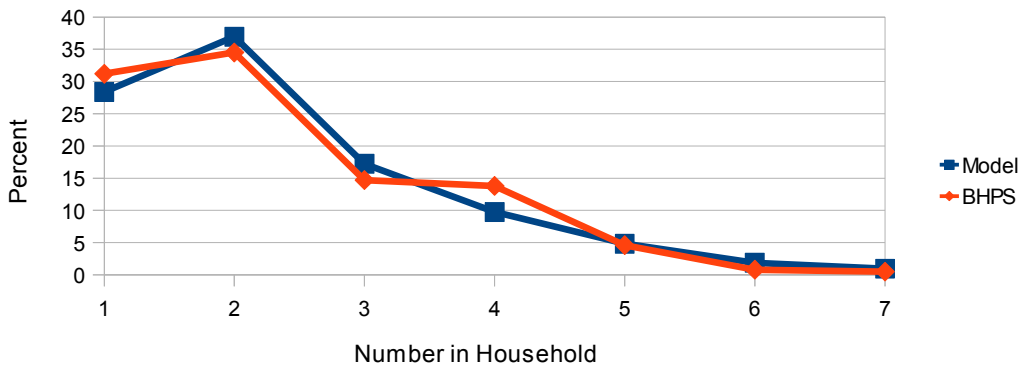


Figure 7: Number in Household Compared Against BHPS 2008

4.6.2.7 BHPS Longitudinal Validation

A significant part of the work of developing the dynamic microsimulation model was in obtaining transition probabilities for births, deaths, partnership formation and dissolution. By comparing the proportion of individuals projected to be in various age

bands and marital status categories, with observed values from the BHPS, an indication of their reliability can be obtained. This is done initially at the individual-level because the transition probabilities were estimated using individual-level data. Next, this is done at the household-level because expenditure patterns will be modelled at that level.

4.6.2.8 BHPS Individual-Level Projections

In these simulations, the BHPS population as at 1991, weighted by household weight (AHHWGHT), was loaded into the model. A series of 10 demographic simulations was then run and the projected population compared against the weighted observed population in subsequent waves of the BHPS until 2008. There is some variation between each run due to the probabilistic nature of the model, which decreases with increasing number of agents. 95% confidence intervals were calculated for each of the simulated data points. The calculation of confidence intervals is only valid for a small sample if it is normally distributed. Since the microsimulation is a type of discrete Markov process, as discussed in Chapter 2, the distribution of results from this simulation would be expected to be normally distributed. Also, in order to verify that there are no artefacts in the program that might distort the distribution, a Shapiro-Wilk test for normality (chosen for its applicability to small sample data) was run on the data points for the last year of the simulation. The results are shown in Table 19.

Variable %	Shapiro-Wilk Statistic	df	Significance
Married	0.947	10	0.634
Single	0.940	10	0.558
Cohabiting	0.880	10	0.130
Separated	0.932	10	0.464
Widowed	0.965	10	0.841
Divorced	0.944	10	0.602

Table 19: Shapiro-Wilk Test for Normality

Table 19 shows that the significance level for all of these tests is well above the threshold of 0.05 which would indicate that the null hypothesis of normal distribution could be rejected. As a consequence, these data are consistent with a normal distribution and confidence intervals can be calculated with the small sample.

The results for partnership status and age are shown in Figures 8 to 10. The solid line represents observed BHPS data. The dotted line is the model projection.

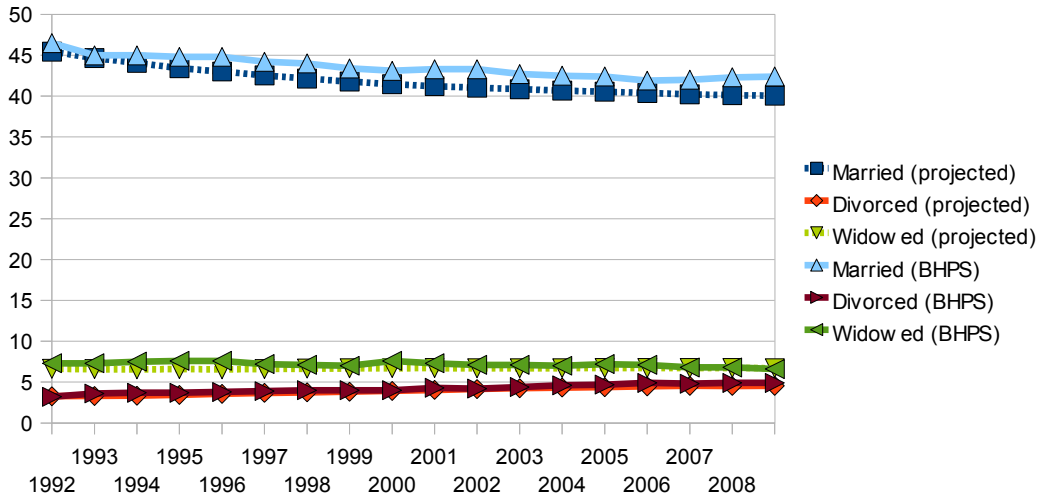


Figure 8: Marital Status (married, divorced, widowed)

The largest interval was calculated to be 0.3% which is less than the thickness of the lines used to show each data series. Hence, it is not shown. The mean 95% confidence interval was 0.15%. This indicates that stochastic variation is negligible for the purposes of this model. The population consisted of approximately 5,500 households and 13,600 individuals. There is a small variation in the initial population compared with the BHPS due to the weighting process.

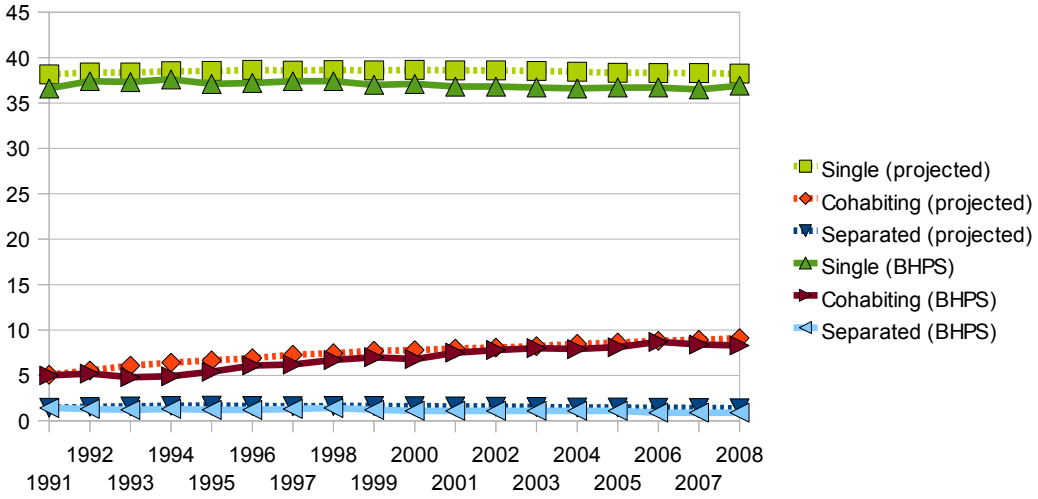


Figure 9: Marital Status (single, cohabiting, separated)

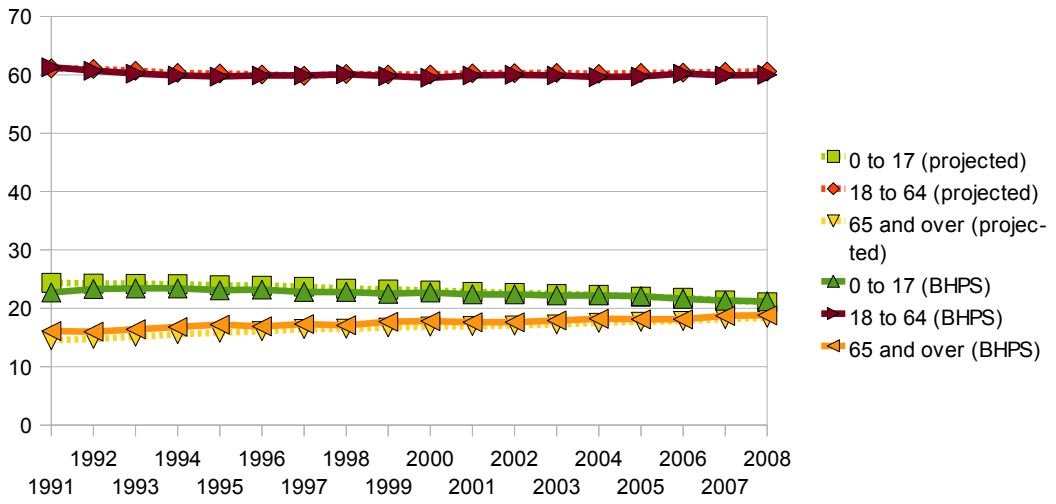


Figure 10: Age Bands

4.6.2.9 Household Level Projections

The demographic modules developed earlier in the project enable the characteristics of a population to be projected over time. In particular, the individual-level processes of birth, death, leaving home, partnership formation and dissolution operate together to determine future household-level characteristics such as the number of people in the household and how they relate to each other. This thesis deals with expenditure at the household level so it will be useful for the purpose of comparison with BHPS data to find a way to describe or characterise the nature of households and project them over time. One way this can be done is to define a set of household types. These were chosen to reflect particular groups of households that might be of interest. A limitless range of types could be chosen but for the purpose of this validation, households corresponding to 'singleton households' (retired and non-retired separately), 'couples with no children', 'families with children', 'lone parent' and 'all other households' were formed. The definition of each type also had to be defined in an objective form so that they could be represented unambiguously in program code and additionally,

the types were selected to correspond to some of those used in the BHPS so that the results could be compared against them.

The variable wHHTYPE in record wHHRESP is available for all waves of the BHPS and has values of:

-9	Missing
1	Single Non-Elderly
2	Single Elderly
3	Couple: No Children
4	Couple: Dependent Children
5	Couple: Non-dependent Children
6	Lone Parent: Dependent Children
7	Lone Parent: Non-Dependent Children
8	2+ Unrelated Adults
9	Other Households

The number of categories was reduced to six for simplicity and are defined as follows:

Definition	**Description**	**Abbreviation**
One person non-pensionable	singleton household	1 non Pen
One person pensionable	singleton household	1 Pen
Two adults	couple with no children	Couple
Two adults + 1 or more children	family with children	Family
One adult + 1 or more children	single parent	Single Par
All other households	other	Other

(An individual is considered an adult if they are aged 16 or over and a child if under 16. Males aged 65 and over are considered pensionable as are females aged 60 or over)

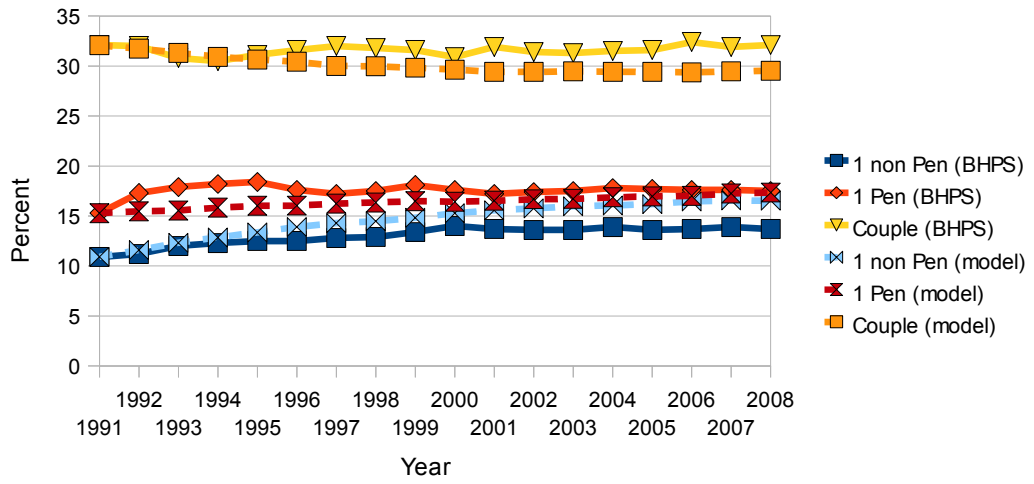


Figure 11: Household Types (single non-pensioners, single pensioners, couples)

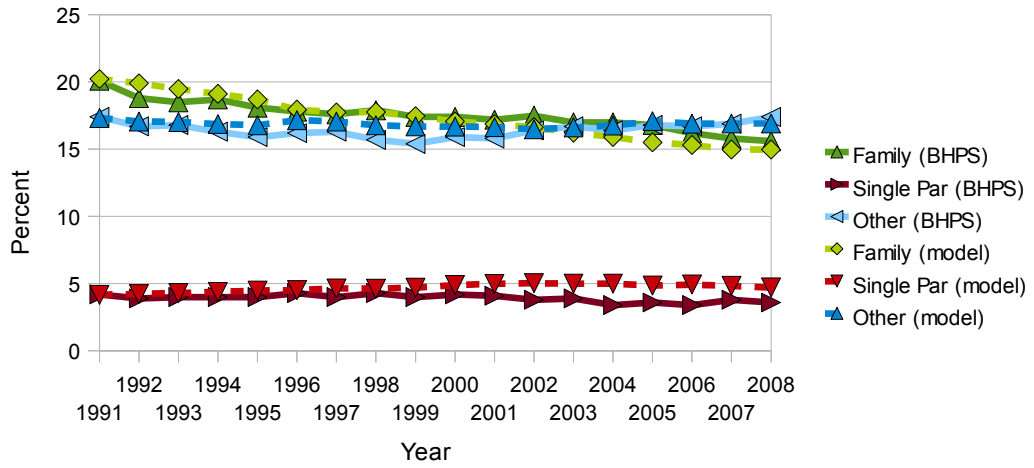


Figure 12: Household Types (family, single parent, other)

4.6.2.10 ONS Cross-sectional Validation

This section of the validation procedure tests whether the model produces plausible long-term projections by comparing its output against projections made by the ONS. The model can be adjusted to compensate for ONS assumptions by varying the transition probabilities.

The next graph shows the projected age distribution against ONS projections for 2030.

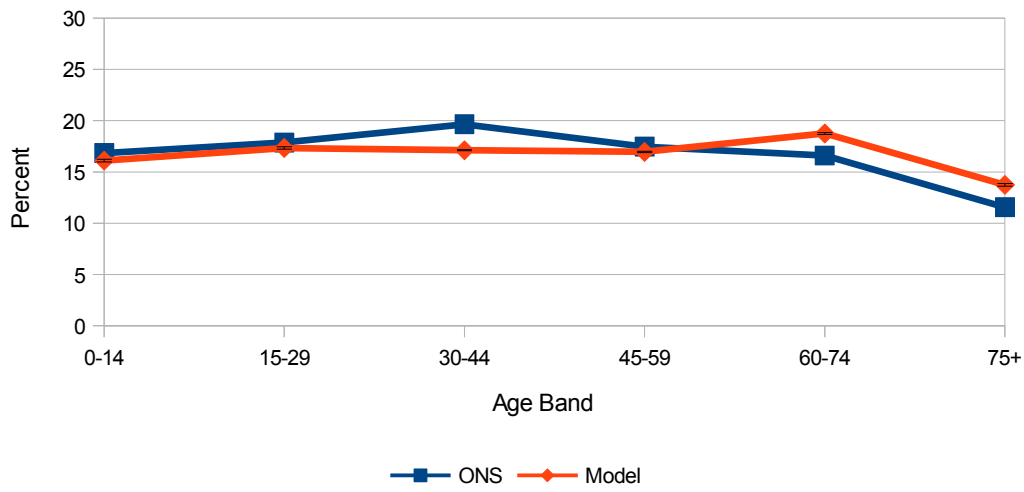


Figure 13: Projected Age Distribution of UK Population in 2031

4.6.2.11 Alignment

The transition probabilities for the demographic model were estimated using BHPS data from the period 1991 to 2006. If the population is projected on this basis, the birth and death rates will reflect the situation that pertained during this time. However, some of the uncertainty surrounding population ageing is due to the fact that birth and death rates are expected to change. The ONS provides projections for mortality and fertility rates in the UK. These indicate that cohort life expectancy, which adjusts for anticipated influences on life span, is projected to rise from 86.8 for males and 90.8 for females in 1991 to 94.2 for males and 97.2 for females by 2035 (ONS, 2011a). The Total Fertility Rate (TFR), which is the expected number of births per female, changed from 1.82 in 1991 to projected values of slightly over two in

2012/13 before falling to 1.84 by 2031.

These projected changes are allowed for in the demographic model by varying the annual transition probabilities for births and deaths by an amount which causes the observed rates from the model to approximate the number of projected births and deaths from the ONS. Figure 14 shows a comparison between the birth rate projected by the ONS and that obtained by using the aligned demographic model. The error bars show the 95% confidence interval of five simulations.

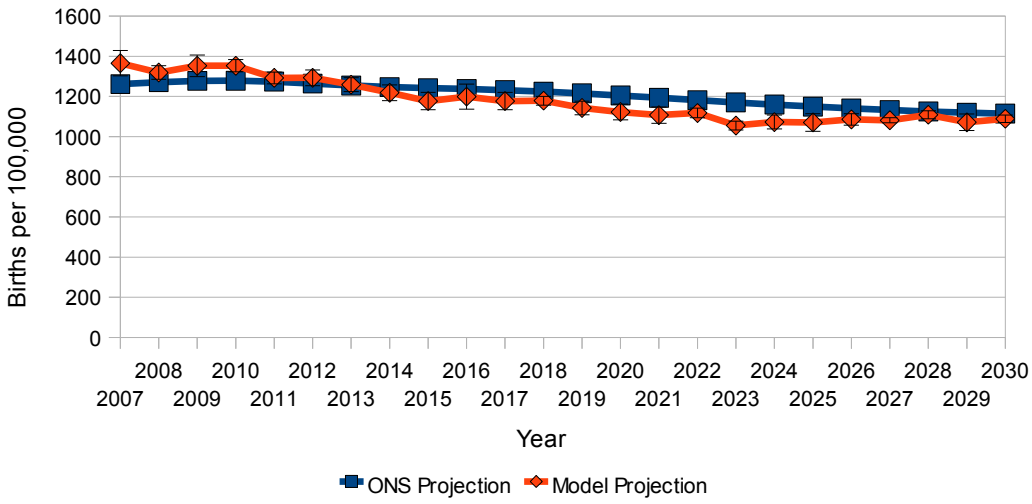


Figure 14: ONS and Model Projected Birth Rates

Figure 15 below, shows ONS death rate projections against those produced from the aligned model with 95% confidence intervals from five simulations.

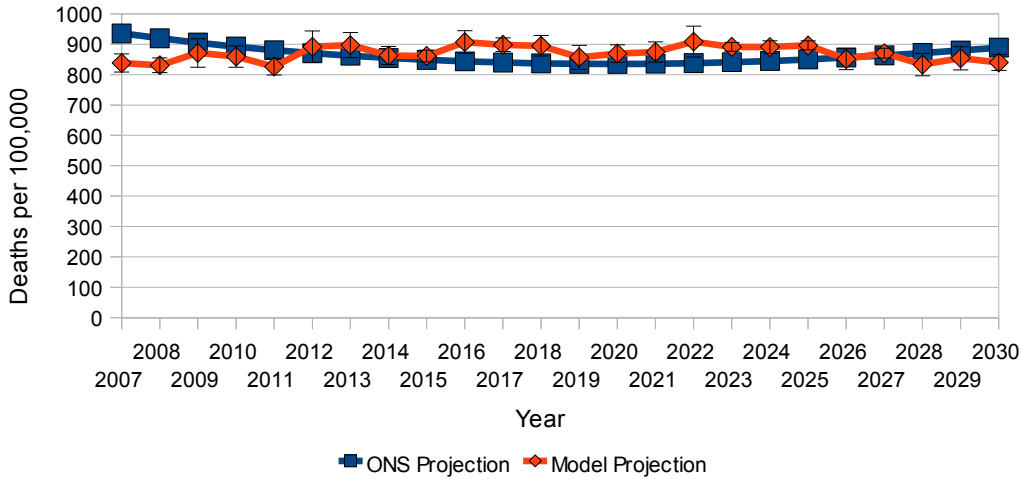


Figure 15: ONS and Model Projected Death Rates

It can be seen from Figure 15 that the population obtained using the aligned demographic model approximates ONS estimates to within 2.5%. It would be quite straightforward to simply select cases at random for births and deaths until the count reaches the value of the ONS projection. If this was done, the age distribution would match the projections exactly. However, this would override the contribution of variables such as partnership status, number of children and age of the oldest child that were included in the transition equations. While the count of births and deaths would match ONS projections, the make-up of the population would be altered. The arrangement used here is simple to implement and strikes a balance between approximating ONS assumptions (which are themselves projections) and exploiting the information provided by the microsimulation model. It will be shown in section 6.4.4.3 below, that the final expenditure projections are relatively insensitive to these assumptions.

4.6.2.12 Impact Validation

In APPSIM, impact validation is concerned with developing confidence that the model provides an accurate simulation of the effect of a policy change. Some progress on this can be made by using aggregate totals generated during the simulation to check for internal consistency. Further checks can be done by duplicating the results of previous macroeconomic models. The APPSIM team found that impact validation is particularly challenging in applications that have not been anticipated by the developers. The extensive use of alignment to ensure that the model reproduces external estimates, adds to confidence in using the model and allows exploration of counterfactual scenarios.

Impact validation in this thesis will be concerned with checking that the effect of demographic change on expenditure patterns is represented accurately using the random assignment scheme. This will be done firstly by checking that the model can reproduce a known function. Then by validating simple changes, that can be checked arithmetically. Finally, more complex scenarios will be tested against results obtained in previous literature where similar questions have been tackled. This part of the validation will be left until Chapter 6 when the demographic model is combined with a random assignment scheme to project the effect of demographic change on household expenditure patterns.

4.7 Comments on the Validation

The validation of Tyche followed the stages described in Morrison (2008). These

ranged from checking data and parameters to testing each module both individually and in combination. Some of the projections were in quite close agreement with observed data. The simulations of partnership type and age bands fall into this category, with the least accurate projection lying about 2% away from the observed value after 17 years. Projections at the household-level were less accurate than those at the individual-level. While some were close to observed values, the model underestimated the proportion of 'families' by 3% and overestimated the proportion of 'other' households by 5% after 17 years. This seems to indicate that while the individual-level transition probabilities are quite accurate, as discussed above, additional error is introduced by limiting the representation of processes that drive household formation to births, deaths, leaving home, partnership formation and dissolution as is usual in microsimulation modelling. Further research would be particularly useful in areas such as leaving and returning to the parental home, the formation of intergenerational households and multiple family households. The longer term projections of birth and death rates could only be brought into agreement with official projections by alignment. However, it is by no means certain that the ONS figures represent the correct version of the future that will inevitably occur in practice. As Harding et al. (2010) note, comparison with external projections should be treated with caution because they are heavily dependent on the assumptions made.

These results echo some of the findings for APPSIM. While many of the benchmarking tests reproduce observed data quite accurately, there are some examples where this is not the case. The percentage of 20-24 year old s in full-time

employment in the year 2047 is projected to be slightly over 40% by APPSIM, while the treasury projection is just over 60%. APPSIM projects that over 20% of people aged 15 to 24 who were employed part-time in 2004 will be in full-time employment in 2005 (Harding et al., 2010: 56). Observed data from HILDA indicates that the actual figure is less than half this value (Harding et al., 2010: 58). The APPSIM team also find that, in many cases, issues with the model or data only come to light when a new policy question is asked.

A dynamic microsimulation model is highly complex. Even for a high quality model like APPSIM, it is impossible to validate every conceivable output. The validation process described above focussed on a number of key elements such as checking that the transition probabilities, with alignment, give rise to a reasonable population trajectory, that households maintain a coherent form and that the population distribution approximates official projections. The essential question for the validation process is whether the program does what it was built to do. The results indicate that the model is suitable for its intended purpose of producing reasonable and plausible population projections for use with the expenditure module to be developed in the next chapter.

4.8 Projected Demographic Change

With the demographic model in place, it is now possible to generate some initial population projections. The model was run to project how the age structure of the UK population can be expected to change over the period 2006 to 2036. It was found that

the number of *simulated* agents rose from its initial value of 58,627 to 64,392 representing an increase of 9.8%. In 2006, the real UK population, according to the ONS, was 60,587,349 (ONS, 2007a). Dividing the two (60,587,349 / 58,627) gives a ratio between the real and simulated population of 1033.44. This implies that each simulated individual represents slightly over 1033 real individuals and allows the model results to be given in terms of the actual UK population.

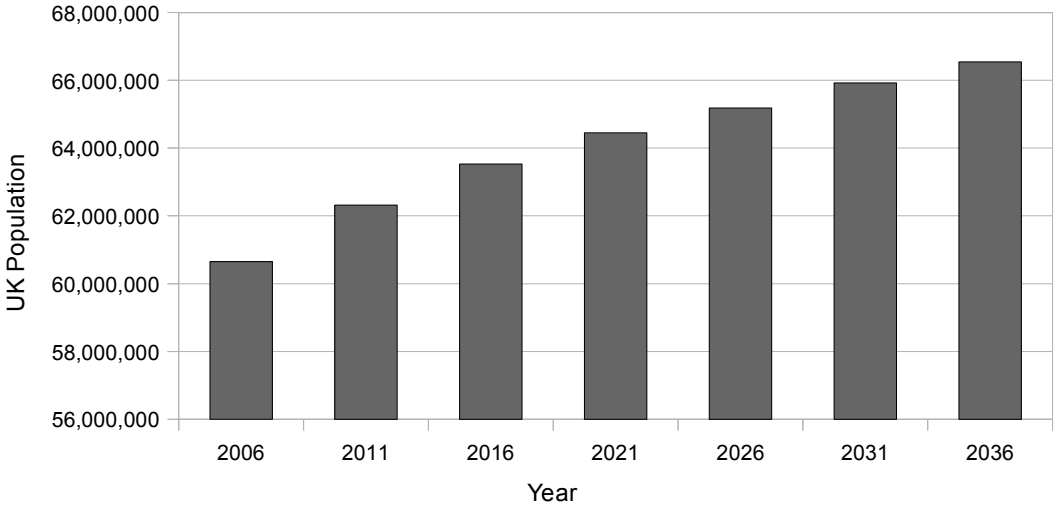


Figure 16: Change in UK Population

One of the most significant features is that the population is expected to rise from 60.6 million to 66.5 million as shown in Figure 16. This is due to natural change alone and takes no account of population migration.

Although the Total Fertility Rate is below 2 for most of this period, the preponderance of women of child baring age in the base population leads to a temporary increase in population size. Figure 17 below, shows that the number of households is projected to

rise by 12.7% from 25,610,710 to 28,858,845.

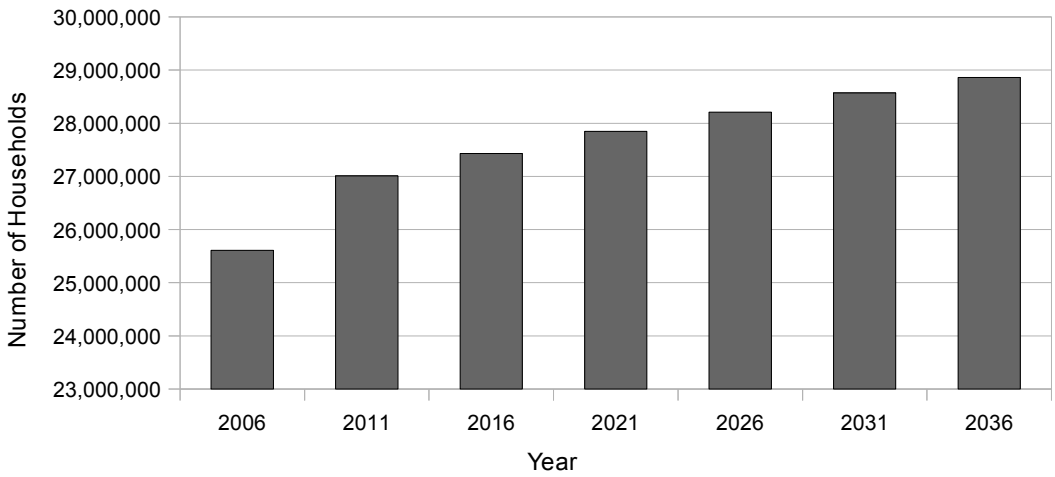


Figure 17: Number of UK Households

Figure 18 below, shows how the age distribution of households changes over time. The age of the oldest person in the household is grouped into ten year intervals and the bars within the age bands show how the number of households in each group is projected to change in five year intervals from 2006 to 2036.

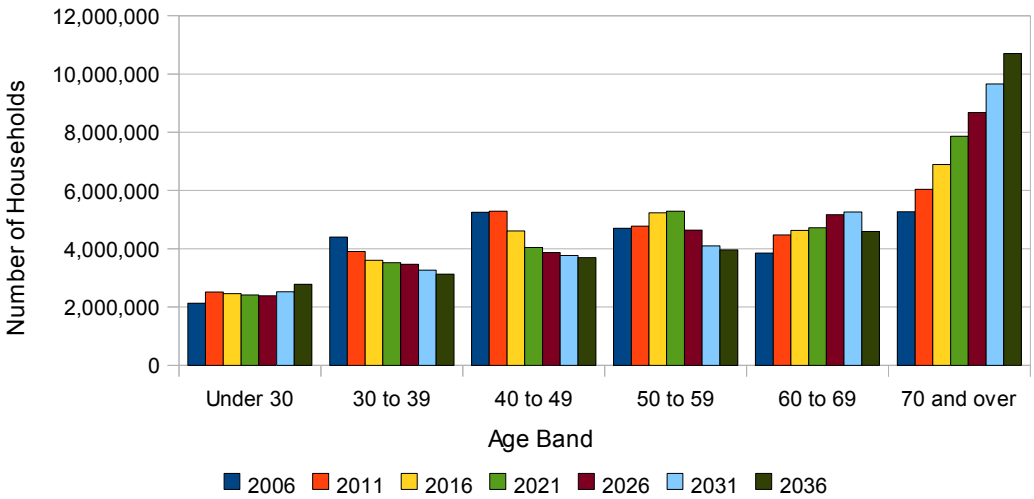


Figure 18: Age Distribution of Households

From Figure 18, it can be seen that much of the increase in the number of households takes place in the '70 and over' age group. The number of households in which the oldest person is over 70 almost doubles while the number headed by a person aged '30 to 39', '40 to 49' and '50 to 59' all decrease. Inspection of the EFS indicates that the older age groups tend to live in households with fewer occupants. The combination of factors of an increase in the number of people aged over 70 and their tendency to live in smaller households leads to an increase in the number of households occupied by one or two people, with three person households also increasing slightly, as shown in Figure 19 below.

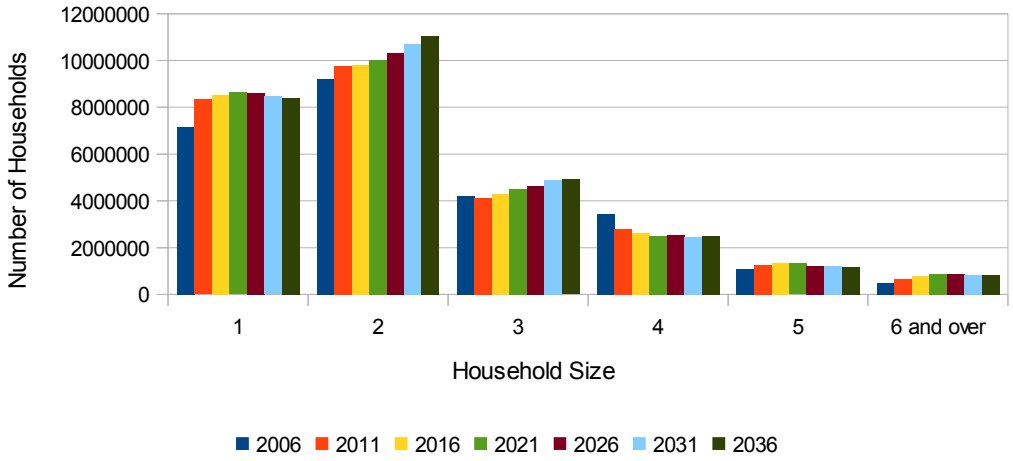


Figure 19: Household Occupancy 2006 to 2036

Hence, the trend towards smaller household size explains why the number of households increases faster than the number of individuals in the population. These initial descriptive results on population change provide a context within which the expenditure projections of Chapter 6 will be assessed.

4.9 Limitations and Further Work

As noted above, a model is an abstraction of certain features from the target system. In the process of developing Tyche, several elements of the real world were left out. This section highlights some of the most significant omissions and the reasons for these decisions.

There is no representation of migration in this model. The only way individuals can enter or leave the simulation is in the mortality or fertility modules. Migration could be implemented by creating or deleting individuals provided there was some method of assigning appropriate characteristics to the new arrivals and selecting which ones would be most likely to leave. However, since the objective of the models described later is to determine the effect of socio-economic changes on the current population, migration is not implemented because it would obscure the effect of intrinsic population ageing which is the object of study for the model.

While spatial relationships are a key element in many microsimulation models, these are not represented here. Although the model operates at the micro-level, the aim is to obtain aggregate results for the whole UK population. In this context, regional variations would be superfluous and so were not implemented. Nevertheless, the micro-level datafile that is available after each simulated year was used as the base population for a model that combined demographic, economic and spatial components (Anderson, De Agostini and Lawson, 2014) which was used to model the effects of austerity measures in the UK.

In a static microsimulation, since the specification of the tax, benefit or pension system is well defined, sampling variation is an important, if not the main source of uncertainty (Goedeme et al., 2013). In the dynamic microsimulation model developed here, there is additional error due to the specification of the demographic system and stochastic variation. Since the demographic modules and method of estimating the equations were done following the approach of previous dynamic microsimulation models, it can be expected that the magnitude and effect of uncertainty and errors is similar to those reported previously. These issues have been documented by, for example, Wolf (2001) and further investigation into the source and effect of these, for this particular model, would largely be a repetition of previous work. However, the random assignment element has not been applied to modelling expenditure before and the effect of stochastic variation in this context is unknown. This will be investigated further in Chapter 6 when the demographic model is combined with the random assignment component, to project household expenditure patterns over time.

4.10 Comparison with LIAM 2

One of the aims of developing Tyche was to evaluate the suitability of NetLogo as a microsimulation platform. The existence of the model demonstrates that NetLogo has sufficient functionality to be used in this role but it does not show whether or in what circumstances it performs better or worse than existing tools. The next section addresses this issue by evaluating Tyche in comparison with LIAM 2.

LIAM 2 is becoming well established as a framework for the development of

dynamic microsimulation models. It originates from LIAM (Life-cycle Income Analysis Model) which has been used to create a range of microsimulation models (O'Donoghue et al., 2009). LIAM 2 is currently under development at the Belgium Federal Planning Bureau (Dekkers, 2011). In common with Tyche, this is a closed population, discrete time, dynamic microsimulation model featuring mortality, fertility, leaving home, partnership formation and dissolution modules.

Dynamic microsimulation models typically operate on a large datasets such as those obtained from a major population survey. They may also be required to project many years into the future and run repeated simulations to model a range of possible scenarios or generate confidence intervals for model projections. Processing speed is therefore an important requirement and the following investigation is considers this first. Next, the exercise compares other features such as usability and flexibility. Table 20 provides a brief overview of the modules associated with each platform.

Function	Tyche	LIAM2
Data Loading	Reads from tab separated text file.	Uses HDF5 file format created by utility for converting from .csv files.
Mortality	Probability based on UK life tables allowing for age and sex.	Aligned to external .csv file.
Fertility	Logistic regression equations for married, cohabiting and unpartnered women. Depends on age, number of children and age of youngest child.	Applies to females aged 15 to 50. Aligned to external file. Creates and initialises new agent.
Partnership Formation	Separate logistic regression equations for marriage and cohabitation depending on age, sex and current partnership status. Partners selected by probability depending on age difference. Alternates randomly between female and male dominated modes.	Applies to people aged 18 to 90. Aligned to external file. Makes list of males and females then matches them according to age, work status, age difference and educational level.
Create New Household	Agents move to unoccupied patch.	Set region at random
Leave Home	Probability depending on age.	Applies to all aged 24 and older not living alone. Starts and initialises new household.
Partnership Dissolution	Separate logistic regression equations for marriage and cohabitation depending on age.	Probability depending on number of children, duration of partnership, age difference and work status. Aligned to external file.
Education	Probability, derived from SAGE data, of completing education in the current year depending on age.	Random selection between levels.
Employment	Students completing education are assigned to employment. Males retire at 65 and females retire at 60.	Set of logistic regression equations involving age, current employment status, marital status. Aligned to external file.
Output	Summary statistics calculated at for each year. Output to charts and .txt file.	HDF5 file with variables for each entity for each time step.

Table 20: Functionality of Tyche and LIAM2

While Tyche and LIAM2 both perform similar functions, in several cases they accomplish these in different ways. One is that Tyche is loaded from a plain text file

while LIAM2 uses a HDF5 (Hierarchical Data Format) file. Tyche implements mortality using transition probabilities, with optional alignment, whereas LIAM2 uses an alignment file for mortality and generally seems to give a higher priority to alignment. Partnership formation in Tyche is implemented separately for marriage and cohabitation, alternating between male and female dominated modes, while LIAM2 combines these into one module using a mate matching algorithm.

4.10.1 Processing Speed Comparison

LIAM2 was downloaded from http://liam2.plan.be/ and installed on the same PC as the version of Tyche against which it was to be compared; in this case, an E-System PC running on a 3.2 GHz Intel Pentium 4 processor with 960 MB of RAM. The downloaded files include a manual that describes in some detail, the operation of each module and also provides instructions on how to use the program. There are several demonstration models and a set of data files containing a synthetic population of 20,200 individuals within 17,700 households.

The data files ‘person.csv’ and ‘household.csv’ were expanded by copying to create a new file containing over 110,000 households. This is comparable with the number of cases in the base dataset of many of the established dynamic microsimulation models listed in (Li and O’Donoghue, 2013). A set of files was created by sampling the first 10,000, 20,000, 30,000 etc. cases until the largest file contained 110,000 households and 151,176 individuals. These were used as the base population in a number of trials to determine how the processing speeds of LIAM2 and Tyche vary with the number

of households modelled. The results are summarised in Figure 20 and represent the mean time of three runs, each of which load the base population and advance it for one simulated year.

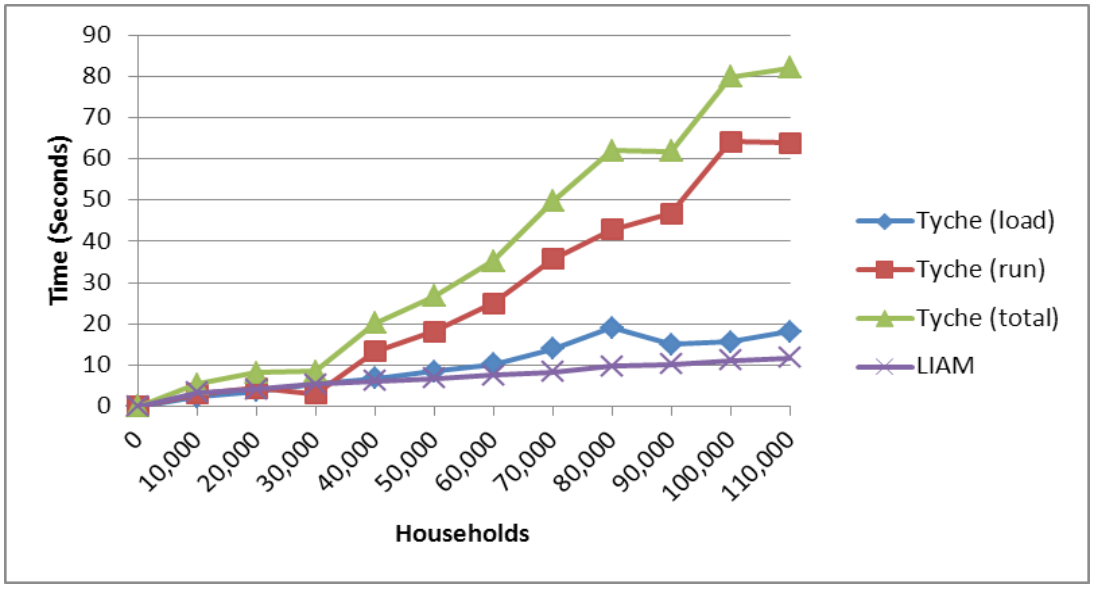


Figure 20: Time to Load and Process 1 Simulated Year

For the largest dataset, containing 110,000 households, the processing time for LIAM 2 is slightly under 12 seconds. Tyche is marginally slower for the datasets containing up to 30,000 households but the processing time for larger populations rises more sharply. In the largest dataset, Tyche took 82 seconds to load and process one year. This is seven times longer than it took using LIAM2.

For the larger datasets, the time taken to load the base dataset text files using Tyche is longer than the combined time to load and run the first year with LIAM 2. The overhead of preparing the HDF5 files used by LIAM 2, which is done using a supplied script file, is repaid in much faster load times. A curious feature of the Tyche

load time is that it took longer to read in the dataset with 80,000 households than the one with 90,000. The computer was rebooted between these runs and it seems that the process of loading data in Tyche is sensitive to the computer's current memory environment. This was not the case with LIAM 2.

One characteristic of NetLogo that is not apparent from Table 20 is that execution time can be very sensitive so the way the program is coded. An earlier version of Tyche took about one minute to load a population of 5000 households and then another minute to advance the population for each simulated year. Two small changes to the code decreased the run time for this model significantly. First, the original load routine operated by reading the household file, then assigning each individual to the appropriate household using the household identification number (wHID) from the BHPS. This was done using the NetLogo primitive 'ONE-OF' which in this case works by creating a set of all patches, then selecting the appropriate one from the list (Railsback and Grimm, 2011). Assigning x y coordinates to the individuals and households at load time, allowed the appropriate household to be found much more quickly. Second, during execution of the simulation, one of the modules required each individual to determine how many dependent offspring it had. This was done by a count routine which again involved polling all agents. Run times were reduced by confining the search to the current household. This could also have been done by keeping a running total of the number of dependants within each agent. In general, code that involves looping through all agents and patches will be slow for large datasets. Usually, as was the case here, there are ways to rewrite the code to be more

efficient. The NetLogo 'profiler extension', which provides timings for user-defined procedures, was useful in locating which sections of code were degrading performance the most. It is quite possible that further improvements in processing speed might be attainable by further optimisation of the code.

4.10.2 Usability and Flexibility

Once the process of installing LIAM 2 and setting up its input files was completed, executing simulations was a straightforward process. The YAML markup code is accessible using a customised Notepad++Portable editor and can be run by pressing F6. The output appears in a console window below the programming window, as shown in Figure 21.

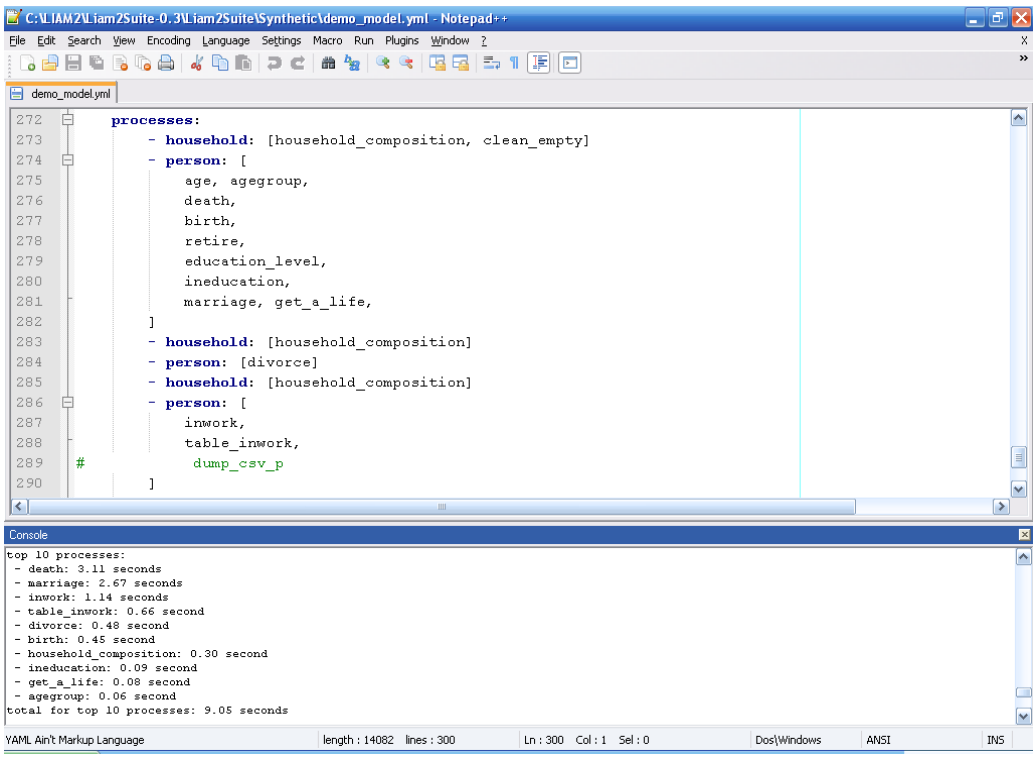


Figure 21: LIAM2 User Interface

It can be seen by comparing the user interface of LIAM 2 above, with that of Tyche in Figures 22 and 23, that NetLogo supports a range of GUI features such as graphs, buttons, sliders input and output boxes, that are not implemented in the LIAM 2 framework.

Figure 22: Tyche User Interface

Also, the world grid facilitates the representation of spatial relationships among the microsimulation units and provides ‘probes’ to dynamically monitor the progress of individual units as shown in Figure 23 below.

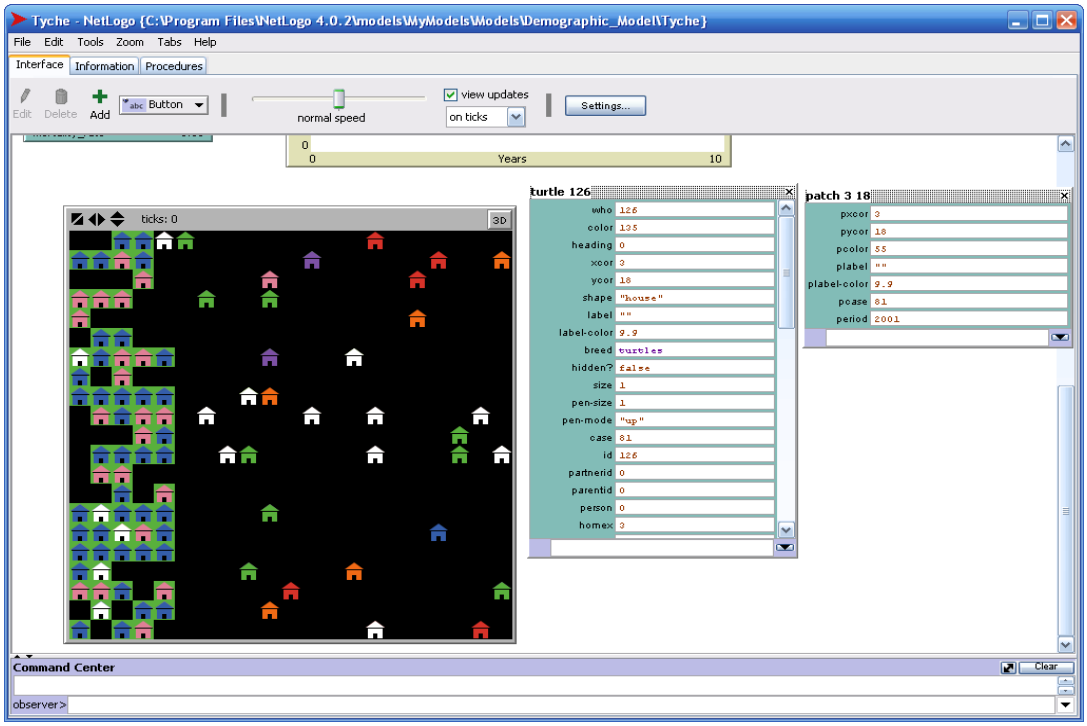


Figure 23: NetLogo GUI Features

In addition, NetLogo has a 'Behaviour Space' feature which allows multiple model runs to be executed and logged. The world grid can be viewed in 3D mode and there is also an option to save the model as an applet and run it in a web page over the internet.

4.10.3 Discussion of the Results

LIAM 2 is a purpose built dynamic microsimulation framework. It has sophisticated alignment facilities and the code is not tied to one particular model. The demonstration version reviewed here, ran about seven times faster than Tyche. Also, LIAM 2 has an active user base within the microsimulation community. While NetLogo was slower than LIAM 2, its ability to load and process a dataset of 30,000 cases in under 10 seconds will be acceptable for many applications. Also, as the

processing speed of computers continues to increase, this disadvantage may diminish in importance in the future.

The NetLogo GUI provides access to some features which are not normally available in existing microsimulation models. One of these is the visual representation of households and their occupants which makes the concept of intra-household linkages intuitive. The world grid may be useful in spatial microsimulation modelling where it is important to consider the physical relationship between entities. Graphical Information System (GIS) information can be read and displayed using NetLogo. Finally, the ability to change parameters as the program is running and view the response in real time via graphs on the user interface makes it possible to explore a range of scenarios without having to spend time transferring to another package to analyse the data.

4.11 Conclusion

This chapter has begun to answer the question of the suitability of NetLogo as a platform for developing a dynamic microsimulation model. The approach taken was to design, construct and validate a program, using NetLogo, that performs all the main functions of a dynamic microsimulation model. The high-level scripting language and programming environment were found to make the development process quite straightforward so that no significant problems were encountered. While the processing speed of Tyche was slower than that of LIAM2, as computer processing speeds increase, this may become less of a limitation in the future. Also,

the use of NetLogo brought with it a range of visual features that are not usually implemented by existing microsimulation models. This provides scope for further development particularly in spatial microsimulation.

In Chapter 2, it was observed that the options available for developing dynamic microsimulation models were: native code (e.g. C++), database programming (UMDBS), a pre-processor (Modgen), a framework written in Python, accessible through YAML (LIAM2) and a 'framework and library' ABM (JAS). The process of developing the model described in this chapter demonstrates that NetLogo can be regarded as a viable approach to microsimulation modelling and so adds a new class of method, (script based ABM toolkit) to the existing toolset.

Chapter 5: Random Assignment Income Model

5.1 Introduction

In Chapter 3, it was argued that established methods of modelling expenditure were restrictive in two important areas. One was that since the number of parameters to estimate rises with the number of goods modelled, data limitations mean that only a relatively small number of goods can be represented. The second was that using dummy variables to describe household characteristics makes it necessary to group households into categories which negates the defining characteristic of microsimulation, which is to work with the individual cases. It was then suggested that a random assignment approach, based on copying expenditure patterns from a similar donor, might provide a way to avoid these difficulties.

This chapter examines random assignment in more detail by means of an example application to model the effect of changes in household income on expenditure patterns. It takes data from the UK Expenditure and Food Survey (EFS) and uses a NetLogo model to project expenditure patterns as household income levels are changed exogenously. Population demographics are kept constant to isolate the income effect. The model validation process is described and results are presented for the whole population, families with young children, pensioners, and by income quintiles. The chapter concludes with a discussion of how random assignment relates to other matching and imputation methods.

5.2 Random Assignment

Klevmarken (1997) describes how what he calls a random assignment scheme can be used as a method of projecting a variable over time. In his example, data is available on a set of incomes for two consecutive years and it is desired to project the income distribution for the following year. Table 21 shows a hypothetical dataset with 10 cases.

Case Number	Income in Year 1	Income in Year 2
1	120	130
2	109	119
3	90	100
4	100	110
5	70	80
6	80	90
7	130	140
8	100	110
9	90	100
10	110	120
Mean	**100**	**110**

Table 21: Initial Dataset

The process begins by defining a distance metric between a donor's income in year 1 and a receiver's income in year 2. In this example, it is simply the arithmetical difference between the two. The essence of the procedure is that the income for each case in year 3 is obtained by finding a donor whose income in year 1 is similar to the current case's income in year 2. The receiver's income in year 3 is then assigned to be what the donor's income subsequently became in year 2, as shown in Table 22.

Case Number	Income in Year 1	Income in Year 2	Donor Case Number	Donor Income in Year 1	Donor Income in Year 2
1	120	130	7	130	140
2	109	119	1	120	130
3	90	100	4	100	110
4	100	110	2	110	120
5	70	80	6	80	90
6	80	90	3	90	100
7	130	140	7	130	140
8	100	110	2	110	120
9	91	101	4	100	110
10	110	120	1	120	130
Mean	**100**	**110**	**Mean**	**109**	**119**

Table 22: Donor Case Assignment

These values were obtained by stepping through each case in turn. Beginning with Case 1, this has an income, in year 2, of 130 units. In year 1, case 7 had a matching income of 130 so it becomes the donor for case 1. Case 2 has an income of 119 in year 2. Case 1 is the closest match with an income of 120 in year 1 so case 1 becomes the donor for case 2. This process continues until all cases have been assigned a donor.

It can be seen that Case 7 requires a donor whose income in the first year is 140. However, since Case 7 already has the highest income, there is no suitable donor and in this situation it is necessary to devise an alternative method of assigning values. In this simple example, the next lower income case is used which is case 7 itself. When all matches have been completed, the donor's income of year 2 becomes the recipient's income in year 3, as shown in Table 23.

Case Number	Income in Year 1	Income in Year 2	Income in Year 3
1	120	130	140
2	110	120	130
3	90	100	110
4	100	110	120
5	70	80	90
6	80	90	100
7	130	140	140
8	100	110	120
9	90	100	110
10	110	120	130
Mean	**100**	**110**	**119**

Table 23: Completed Income Projection

Table 23 shows that the procedure has, for the most part, correctly implemented the implied rule that the income next year is 10 units higher than it was this year. In this example, it is easy to deduce what the rule should be, however the power of this method is that, since the rule remains implicit, it is possible represent an arbitrarily complex system of rules with no additional complexity to the model. This means that a system can be modelled without first having to discover precisely how it works. In modelling aspects of human behaviour, such as consumption, where the rules are complex and may be impossible to discover, this is an important advantage.

The method operates by applying last year's change to project what will happen next year. This is a limitation in that, if incomes changed by a different amount each year, it would not be predictable using this approach. Also, there is an anomaly in that the

mean income in year 3 is 119 and not the 120 that would be expected. This arises from there being no suitable donor for Case 7 and illustrates one of the limitations noted by Klevmarken (1997) which is that it is not possible to predict beyond the range of values already present in the original sample. Nevertheless, the method is flexible enough to allow information to be incorporated from another source. For instance, a regression model could be used to obtain a plausible value for Case 7 or there may be another dataset from which the value could be imputed.

This scheme can quite readily be turned into a method for modelling household expenditure. As an example, there could be a situation where there is a change in household circumstances, such as a new person moves in. It may be expected that spending on some items like food would increase while others such as rent or mortgage would not be affected. The purpose of the microsimulation model is to determine which expenditure categories would change and by how much. This can be done by locating a household within the dataset that already has a composition that is as similar as possible to the new household and copying its expenditure pattern. The matching variables would include demographic characteristics and financial variables such as income. These variables would be chosen because they are thought to have some correlation with expenditure patterns and they play a role corresponding to the independent variables in regression modelling. It is assumed that whatever changes the household in question will make as a result of the new member arriving, would already have been made by the donor household so this behavioural response will be embedded or encoded within the expenditure pattern that is copied. Unlike

Klevmarken's method, the scheme used here is based on a cross-sectional input file, deriving its projections from the heterogeneity of cases, as opposed to a longitudinal dataset which contains variation over time. However, the underlying principle of matching and imputation is essentially the same.

The next section describes a simple implementation of a random assignment scheme to model the effect of changes in income on household expenditure patterns. It first details how the model is constructed. Then it discusses a means of validation to check that the results are plausible. Finally, some disaggregated results are provided to demonstrate how the method can be used to study the way in which different sections of the population react to changes in household income.

5.3 An Application of Random Assignment to Model the Effect of Changes in Income on Household Expenditure Patterns

5.3.1 Level of analysis

Economic modelling is often carried out at the individual level. This makes sense because it is the individual who makes decisions and has some agency regarding their consumption behaviour. However, it is possible for individuals to have no income of their own and yet spend money on a range of items. This is explainable by intra-household allocation of resources and at the individual level, this would have to be represented in the model. Working at the household level encapsulates intra-household allocations implicitly, in observed spending patterns and so simplifies the specification of the model.

5.3.2 Data Source

In order to investigate the relationship between household income and expenditure, keeping demographic characteristics constant, it is necessary to have some information on expenditure patterns that can be linked to household parameters such as the number of people in the household, their ages etc. In the UK, the Expenditure and Food Survey (EFS) provides data on around 2000 spending categories and includes a set of demographic variables describing household characteristics. This makes it suitable for use as the base data set for the model and avoids the need to combine data from more than one source.

The EFS is an annual cross-sectional survey that collects detailed information on household spending obtained from respondents keeping a diary of all spending over a two-week period, combined with retrospective interviews to cover large, occasionally purchased items. Its sample size is around 6,000 households containing over 10,000 individuals. Household and individual-level weights are provided so that the survey sample is representative of the UK population. The illustrative model described below restricts itself to the 12 high-level expenditure groups defined in the EFS, which correspond to the Classification of Individual Consumption by Purpose (COICOP) categories (UN, 2011). Table 24 provides a brief summary of each type and some notes on what is included.

Variable Name	EFS Household Expenditure Category	Notes
P601t	Food & non-alcoholic Drinks	
P602t	Alcohol Tobacco & Narcotics	Alcohol to be consumed at home
P603t	Clothing & Footwear	
P604t	Housing Fuel & Power	Includes rent, maintenance of household, water and fuel bills. Excludes mortgage costs.
P605t	Household Furnishings & Equipment	Includes carpets, curtains, household appliances, utensils and tools
P606t	Health	Prescriptions, glasses, dentist fees but not medical insurance
P607t	Transport	Purchase of vehicles, fuel, vehicle maintenance but not insurance
P608t	Communications	Mobile and fixed line telephone, postage but not internet subscription
P609t	Recreation & Culture	Television, computers, CDs, boats, caravans, pets, sports, holidays
P610t	Education	Course fees, school trips
P611t	Restaurants & Hotels	Includes takeaways, alcohol consumed outside the home and school meals
P612t	Miscellaneous	Includes insurance, jewellery, child care, fees, moving expenses

Table 24: Primary EFS Expenditure Categories

5.3.3 Implementation

The program was developed in the form of a simulation using NetLogo. An overview of the program algorithm is shown below depicting the main stages of the simulation.

```
load cross-sectional household data file from the EFS
for each simulated year
    for each household
        increase income by chosen percentage
    for each household
        locate a donor household that has a similar composition
            and previous income, similar to the current household's new
                                                                income
```

copy expenditure pattern from donor
calculate new aggregated expenditures for categories of interest

5.3.4 Matching Criteria

One of the essential features of this algorithm is to locate a similar household. There are many ways this can be done and the selection of which particular method to use will depend on the application and available data. O'Hare (2000) describes several approaches for matching similar cases. These include minimum distance matching, where the donor case is chosen due to its closeness on one or more parameters. A common distance function is the normalised Euclidean distance $\sqrt{\sum(x_i - y_i)^2}$ where x_i and y_i are the parameters for case i. This can be extended for as many parameters as required. One of the advantages of minimum distance matching is that it can be used on continuous data. If the data are categorical, matching can take place by random selection of a donor case within the same class. These methods can be combined in matching by minimum distance within classes. A further option that could be explored is to use cluster analysis to define equivalence classes and select the donor case from within the appropriate cluster. In the simulation described here, matching takes place using household income, the number of people in the household and a derived variable known as household type. In O'Hare's classification, this is minimum distance within classes.

The model operates on the principal that whenever a household experiences a change in income, the expenditure pattern is copied from a similar household that has already had the new income and has had time to adapt. As a result, income, by definition, will

be one of the matching criteria. It is represented as if it were a continuous variable so there are no income bands. In many cases, there is not an exact match for a particular income so it is necessary to find the closest alternative. In this implementation, for half of the cases (selected by a random draw), the household with the next higher income is selected as the donor. In the other half, the next lower income household is used. If more than one household satisfies the matching criteria, one of them is selected at random to be the donor.

A natural approach to performing the match would be to always choose the closest case, by whatever metric was being applied. If this was done, the same donor and recipient would be paired in each simulated year and there would be a risk that chance initial combinations of households would be amplified over time leading to a distortion of the distribution of cases. Alternating donor cases was found to reduce this effect. Another advantage of the introduction of the stochastic element is that different instances of the simulation give slightly different results depending on the particular donor cases that happen to be selected during a simulation. In consequence, it is possible to obtain a measure of the effect of the selection of donor cases by constructing a statistical distribution over a number of runs, which would not be possible in a single instance of the simulation or if all simulations gave the same result. This only captures the effect of case selection but as some of the examples in Chapter 6 indicate, the effect is considerable for some budget categories.

The next matching variable is the number of persons in the household i.e. household

size. It is clear, from previous research (OECD, 2012), that this variable is related to household expenditure; the more people in the household, the greater its expenditure is likely to be. However, its effect in the context of this model is more subtle. If the only matching criterion was income, then selected donor households would be more likely to be those that have a greater number of occupants, because larger households tend to have higher incomes. Since households are unlikely to adapt to increasing incomes by taking in new members, this effect was removed from the model by including the number of occupants in the matching criteria.

A third matching variable will be household type, with the same six categories that were used in the demographic model described in the previous chapter.

Type	Description	Definition
1	One person non-pensioner	one person below pensionable age
2	One person pensioner	one person of pensionable age
3	Couple with no children	two adults
4	Family with children	two adults + 1 or more children
5	Single parent	one adult + 1 or more children
6	Other	all other households

If there are no appropriate donors (as is the case for households at the ends of the income distribution), all expenditures are changed in proportion to the desired income change and no copying is done for that household in the current cycle.

In order to simulate the effect of changing household incomes, a scenario is run where all households receive a fixed percentage change in income. Processing continues until all households have had their income changed or when the mean

income of the population has been changed by the target amount, whichever comes first. Details of the random assignment procedure are shown below, beginning from the point where all households have been set a desired or target percentage increase in income.

```
for each household
        while mean population income <= target population income
                select uniformly distributed random number from 0 to 1
                if number <= 0.5 (copy from next higher income household)
                        if any other households with (income >= target)
                                and (increase <= limit)
                                and same type and number in household
                                        store donor expenditure pattern
                        else
                                change all spending categories in proportion to income
                                                                                change
                if number > 0.5 (copy from next lower income household)
                        if any other households with (income <= target)
                                and (increase <= limit) and same type
                                and number in household
                                        store donor expenditure pattern
                        else
                                change all spending in proportion to income change
for each household
        update expenditure patterns from stored donor
```

The source code and further details of this model can be found in the OpenABM model library: http://www.openabm.org/model/3918/version/1/view

5.3.5. Running the Model

In order to run a simulation, it is first necessary to set up some user defined parameters. These are the number of cases to read from the input file, the number of cycles of random assignment to run and the percentage change in real household income per cycle. It is possible to vary the rate of income change to represent a particular scenario either via a slider on the user interface or in the program code. The output appears in graphical form on the screen or optionally can be written to a file for later analysis.

5.3.6 Validation

The purpose of this section is to check the reliability of the random assignment scheme in that the results generated from using this method form an accurate representation of relationships within the data. One of the most effective ways to do this would be to make some projections over time and then test whether they are accurate by comparing them with observed data. This is how the demographic component was tested in Chapter 4. In that case, the results were found to be within a few percent after several years, which seems reasonable given the concerns of Winder (2000) who noted that microsimulation models usually fail to simulate known time-series data. The reason this kind of test was feasible is that although demographic change can be influenced by several factors, such as the delaying of births in times of economic austerity, the effect is relatively slow and marginal. Conversely, economic influences on household spending patterns can have a large effect in a short time. Prices, interest rates, new products, fashion, even the weather, can affect demand

significantly.

The model, as specified above uses changes in income to predict changes in household spending patterns. This seems reasonable because it has been known for some time that a budget constraint affects the way money is spent. Engel's Law (Engel, 1857, 1895) for example, indicates that the budget share for food varies inversely with income. This idea has been extended to other goods by Maynard and Beckman (1946) who used it to predict changes in expenditure patterns as incomes change. Subsequently, the general applicability of this approach to predicting budget shares has been tested, statistically, not by simulation, against observed data by Loeb in (1955). In this study, he investigated changes in the budget share of food, recreation, clothing and transportation that were measured in the 1954 National Income Supplement to the Survey of Current Business. These were compared with the direction of change in real household incomes over the period 1929 to 1953. This was done annually, in five year overlapping bands and over the periods 1929-1938, 1939-1953 and 1929-1953. Loeb found that apart from the 1929 to 1938 period, ‘all the tests reveal considerable nonconformity’ to what would be expected by applying the laws. The reasons he suggests for this are many and varied but essentially revolve around idiosyncrasies concerning the individual commodities and the time period over which the observations were carried out. In the case of food for example, he cites rising food prices, a change in diet, increased food processing and improved merchandising as possible explanations for a period of rising budget shares for food at the same time as real incomes were rising.

Using the model to make unconditional forecasts and comparing them against observed data would be an implementation of Maynard and Beckman's method of modelling budget shares. However, we know from Loeb's results that this would often be unsuccessful due to extraneous factors. It would be possible to add additional influences to the model described in this chapter, such as prices, interest rates etc., one by one, until the model produced a good approximation to past data. However, the influences cannot be assumed to combine in an orderly way so that the results gradually improve. Since they act independently, in different directions, some influences might reduce the errors while others would make them worse. Only when the last element was included would the output become accurate. Then, a model developed in this way would be so over fitted to its test data that it might well be no better at forecasting out of sample than a more parsimonious implementation. In addition, it would be so complex that it would undermine one of the objectives of modelling which is for it to be a selective abstraction from the real world which is simple enough to understand and experiment with. A suggested alternative method of validation known as backcasting or 'BackSim' (Birkin and Malleson, 2014) involves calculating the rate of change in the number of aggregated types of household and projecting backwards in time to compare their numbers against observed values. This was not thought to be feasible in this case because the backward simulation could add another layer of errors to the model and it would still, at some point, have to be compared against real data which includes the idiosyncrasies described earlier.

An inability to demonstrate that the model is able to reproduce past data series does

not imply that it cannot be validated or has no use. The items that should be checked or tested depend on what the model is to be used for. For the models developed in this thesis, the aim is to project how spending patterns would change in response to variations in one or more specific variables, in isolation from other confounding factors. This is in the tradition of microsimulation modelling where it is the effect of a policy change and various counterfactual policy scenarios that are of interest, rather than to anticipate real events some years in the future. Unfortunately, it is often the case that the effect of a policy change or variation in income, as modelled here, is difficult to observe in practice because it is embedded within a range of other influences and some kind of model would have to be used to extract the relevant signal from the extraneous noise of all the other factors. Then, some of the error would be due to the microsimulation model under test and some would be due to the other model and it would not be possible to correctly attribute errors to each one. This is one of the reasons why dynamic microsimulation models in general, are difficult to validate, as was found in Chapter 2.

In the light of these limitations, an alternative approach is used here which is built up in three layers. The first is to demonstrate that the random assignment method, as applied in this research, is underpinned by the Law of Large Numbers (Papoulis, 1984) so that the results will accurately represent the distribution of variables in the donor cases, if the sample is large enough. The second layer of validation is, instead of using real data to test the projections, a synthetic dataset that contains a known relationship between income and expenditure is used to demonstrate that the random

assignment scheme is able to model it correctly. Finally, a microsimulation model is tested using real data to check that it can reproduce a number of stylised facts that have been established in previous research.

5.3.6.1 Theoretical Validation

It can be demonstrated that this use of random assignment is based upon the Law of Large Numbers by constructing a simple example. There are imagined to be two types of household, Type A and Type B. Households within each type are identical with respect to their demographic composition and income but the characteristics of Type A are different to those of Type B. Each individual household has its own idiosyncratic spending pattern which is nonetheless, influenced by its characteristics so that the mean expenditure on a set of goods for Type A households is different from the mean expenditure for Type B households.

If a number of Type B households alter their characteristics to become Type A households, they will each copy the spending pattern of a randomly selected Type A household. It is apparent that the distribution of spending patterns for the recipient households will approach that of the Type A households, provided that a sufficiently large number of households make the transition or if the process is repeated a sufficient number of times.

5.3.6.2 Reproduction of a Known Relationship

The second part of the validation is to show that the random assignment scheme is able to represent correctly, a known relationship within its input dataset. This begins by setting up a dataset where the expenditure for a good maintains a fixed relationship with income. In this example, expenditure for the good is always 10% of income, as shown in Table 25, which lists the first ten cases of the artificial dataset.

Income	**Expenditure = Income / 10**
291.80	29.18
255.11	25.51
215.75	21.58
185.80	18.58
850.08	85.01
258.49	25.85
155.55	15.56
268.72	26.86
157.17	15.72
179.56	17.96

Table 25: Ten households which spend spend 10% of their income on a good

The model described above was run such that UK household incomes rose from their initial mean average level of 654.79 to 1066.78. When this was done, the expenditure projection correctly modelled the rule that expenditure is always 10% of income as shown in Figure 24.

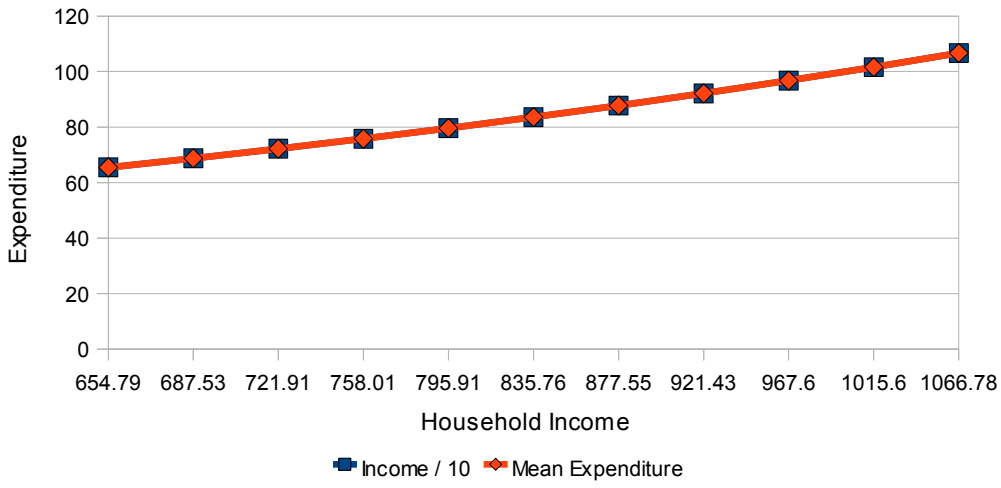


Figure 24: Results from modelling artificial rule

This result is not surprising because the model is simply reading off the correct expenditure for every household. It is also apparent that any relationship whether it be non-linear or highly complex, would be represented correctly without having to code it into the model because the required information is contained within every single household (all households use the same rule).

In reality, the exact value of expenditure is not perfectly determined by income, there are many other factors involved and they can vary from one household to the next so that, however carefully the model is specified, there will still be a certain amount of error. This can be simulated by applying a random disturbance to the expenditure of each household. In this example, the expenditure function of income divided 10, was modified by adding a normally distributed random variable with a mean of 0 and a standard deviation of 100. In cases where the application of the random disturbance

led to a negative expenditure, its value was set to zero. The first 10 cases of the new dataset are shown in Table 26.

Income	Modified Expenditure with Random Disturbance
291.80	28.92
255.11	23.83
215.75	173.50
185.80	56.56
850.08	7.88
258.49	0
155.55	0
268.72	114.14
157.17	84.40
179.56	108.89

Table 26: Input Data with Stochastic Disturbance

The target value for each case's expenditure was calculated using the same algorithm and this was compared with the output of the model in a simulation with a population size of 1,000 households.

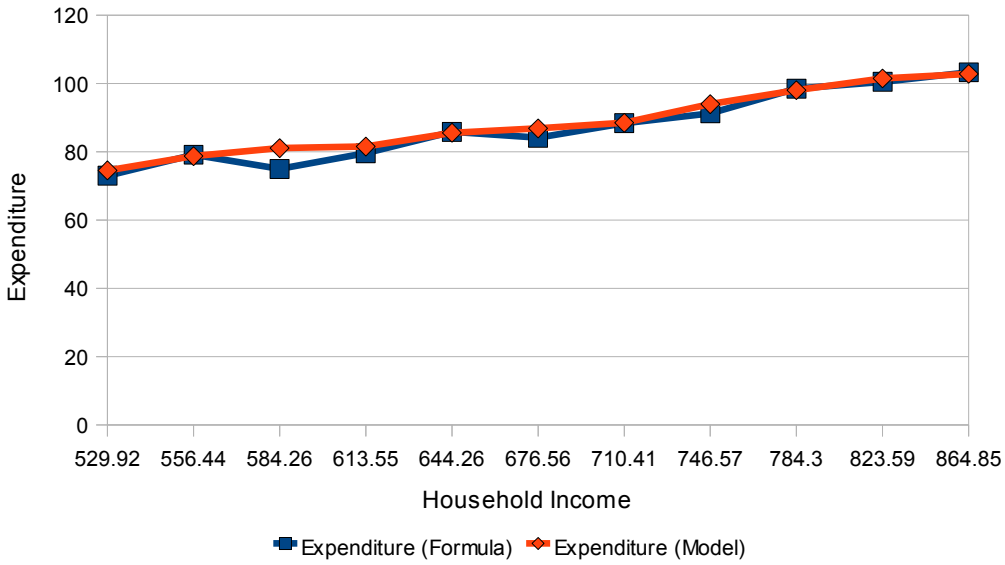


Figure 25: Modelling an Artificial Rule with Random Disturbances

It can be seen from Figure 25 that the modelled function approximates the calculated function quite well. The only difference is due to random variation in the two series. This can be decreased with by using a larger population or by running the simulation a number of times and averaging the results.

5.3.6.3 Stylised Facts

The last piece of validation is to base the model on real data from the EFS and test whether it produces features of the relationship between income and household spending that have been observed in previous research. The relationship between household income and spending is an area that has been studied quite extensively. The model described in this section is not intended, primarily, to add to the voluminous literature on this subject. Rather, a few stylised facts are abstracted from what is known and these are used to test the validity and plausibility of the results

produced by the model.

One of the most obvious features of household spending patterns is that total consumption increases with income. However, as incomes rise, not all of it goes to consumption expenditure; some is saved or invested and some is paid in income tax. This means that, as household incomes rise, total expenditure will increase at a slower rate than income. Aside from total expenditure, a significant amount of research has been done on how the share of expenditure for goods varies with income. As far back as 1857, Engel found that the budget share for food decreases as household income increases (Engel, 1857, 1895). More recently, ONS figures (ONS, 2008) indicated that households in the highest income decile spend a greater proportion of their expenditure on 'transport' and 'education' while spending a smaller proportion on 'housing' and 'food' compared to the lowest income decile.

5.4. Results

5.4.1. Average Spending

When the initial data file, taken from the 2006 EFS, was read in, mean UK household income (before tax and including any benefits) was £655. This was increased in a series of 10 steps (10 cycles of random assignment) until it reached £1058, which represents a 62% increase. During this time, mean total consumption expenditure rose from £478 to £654, which is an increase of 37%. This is in agreement with the first two stylised facts, identified above, that expenditures rise with rising income but at a slower rate.

Although it is the primary output from the simulation, it would not be appropriate to write out the disaggregated micro-level dataset here. The following aggregated summary shows how the population reacts to increases in household income. This can be compared against the stylised facts mentioned above to verify that the model produces the expected results.

Table 28 shows expenditure changes for the whole population. The 'Initial Share' is the proportion of total consumption expenditure spent on each category at the start of the simulation, averaged out over the whole population. The 'Final Share' is the budget share for the item after incomes were inflated as described above. 'Shift in Budget Share' is the difference between the initial and final share. Due to the stochastic nature of the simulation, there is some variation from one run to the next. As a result, it is not possible to know whether the results from a single instance are typical or unusual. Running the simulation a few times will provide information on how much random variation takes place between runs. This can be quantified in a number of ways such as calculating the difference between maximum and minimum values of the output for a particular parameter, obtained from a number of runs or from their standard deviation. It is also possible to calculate confidence intervals for the results and the next column of Table 28 gives the 95% confidence interval of the final budget share for each expenditure category. The last two columns of Table 28 show the upper and lower bounds of the confidence intervals. The results are the mean of five simulations. Confidence intervals are only valid for a small sample if the data is normally distributed. There is good reason to believe that the distribution of

results over multiple simulations is normally distributed because the process generating the data is a discrete Markov Process (Andreassen, 1993). In order to check that this implementation does not violate the conditions under which this belief is true (the events are random and independent of each other), a Shapiro-Wilk test for normality was conducted on the output for the final year of the set of simulations, the results of which are shown in Table 27.

Category	Statistic	df	Significance
Total Expenditure	.957	5	.788
Food	.882	5	.317
Alcohol	.754	5	.032
Clothing	.879	5	.303
Housing	.863	5	.239
Furniture	.881	5	.312
Health	.881	5	.313
Transport	.926	5	.570
Communication	.969	5	.872
Recreation	.861	5	.232
Education	.853	5	.205
Hotels	.975	5	.904
Miscellaneous	.967	5	.854
Other Expenditure	.926	5	.568
Mean Income	.960	5	.806

Table 27: Shapiro-Wilk Test for Normality

In Table 27, for all categories except 'alcohol', the level of significance indicates that it is not appropriate to reject the null hypothesis that the data is consistent with a

normal distribution. In the case of 'alcohol' this is probably due to the results from two of the simulations happening to be identical and so by chance, distorting the distribution. While this gives some indication of the variability of the model, it gives no indication of the accuracy of the model specification.

Table 28 is sorted into descending order of percentage change so that as incomes rise, households devote a greater proportion of their consumption expenditure to the goods near top of the list. These categories can be said to have a greater discretionary element than those lower down.

Expenditure Category	**Initial Share (%)**	**Final Share (%)**	**Shift in Budget Share (%)**	**95% Confidence Interval (+ -) α = 0.05, n = 5**	**Lower CI**	**Upper CI**
Other	20.87	23.07	2.2	0.402	22.67	23.47
Transport	12.97	13.9	0.93	0.196	13.7	14.1
Education	1.49	2.26	0.77	0.152	2.11	2.41
Furniture	6.33	6.49	0.16	0.059	6.43	6.55
Health	1.23	1.35	0.12	0.056	1.29	1.41
Miscellaneous	7.55	7.62	0.07	0.088	7.53	7.71
Clothing	4.86	4.76	-0.1	0.085	4.68	4.85
Hotels	7.92	7.62	-0.3	0.080	7.54	7.7
Communication	2.46	2.09	-0.37	0.023	2.07	2.11
Alcohol	2.33	1.93	-0.4	0.027	1.9	1.96
Recreation	12.25	11.67	-0.58	0.141	11.53	11.81
Housing	9.94	8.93	-1.01	0.078	8.85	9.01
Food	9.81	8.13	-1.68	0.060	8.07	8.19

Table 28: Share of Expenditure with Increasing Income (all households)

Table 28 indicates that the categories of ‘education’ and ‘transport’ receive a greater share of the increased expenditure than the other items. This is consistent with the ONS results mentioned above, which showed that households in the highest income

decile spend more of their income in these categories than those in the lowest income decile.

At the bottom end, 'food' and 'housing', receive a diminishing share of the rising income, which is consistent with what was reported by the ONS. Finally, the budget share for 'food' falls as incomes rise and this is in agreement with Engel's Law.

These results are not surprising. They are merely intended to verify that the model produces the expected results. However, the micro-level cross-sectional output can be analysed in a variety of ways and the next two sub-sections provide further results for selected groups within the population that can be expected to have differing expenditure patterns. The first of these compares 'older households' and 'families with young children' with the population mean household spending level in section 5.4.2. After this, section 5.4.3 gives results disaggregated by income quintile.

5.4.2 Older Households and Families with Young Children

Figure 26 shows changes in spending patterns for two types of household:

a) those with at least one person aged over 65 (older households)

b) households with one or more children under 5 (families with young children).

In each case, incomes were inflated in the same way as in the previous simulation, then spending in pounds-per-week, for each good, was divided by the square root of

the number of people in the household to compensate for household size (Buhmann et al., 1988). Next, the amount spent on each expenditure item was divided by the amount spent on that good by the average for the population so that the results are shown as a percentage of what the typical household spends. The expenditure categories are shown on the x-axis: first for the older households (65+) and then for the family with young children (F). Each individual bar represents spending at each step of the simulation as incomes were increased.

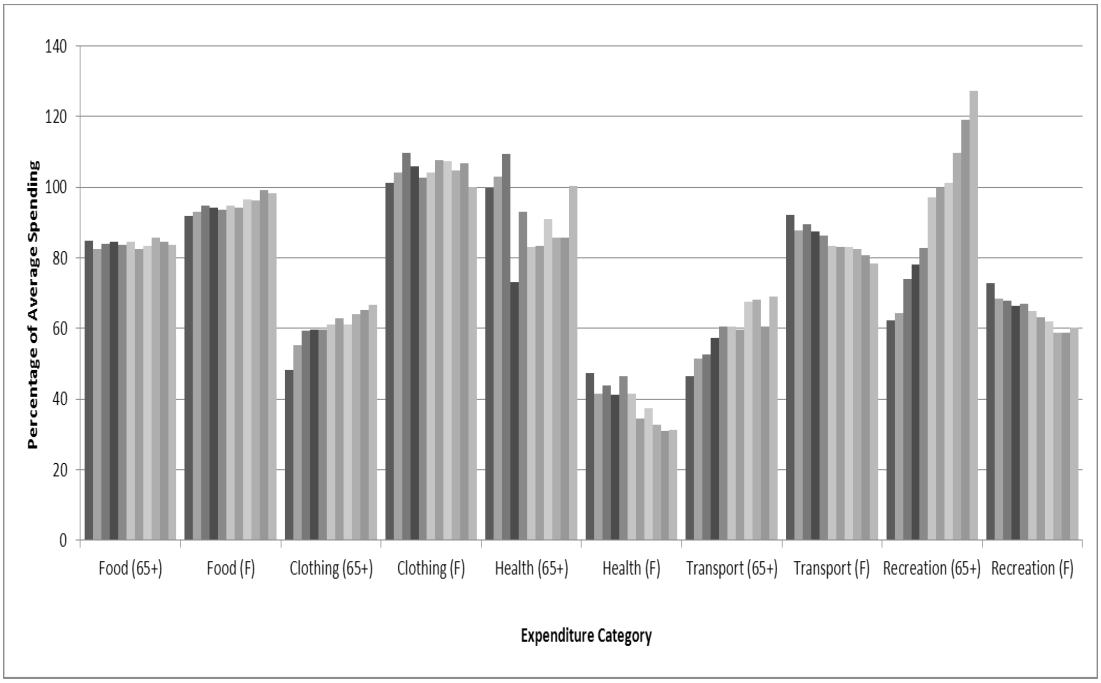


Figure 26: Share of Expenditure (older households and families with young children)

These results were obtained from one simulation so there are no confidence intervals. However, the amount of variation between each cycle of random assignment can be seen in the level of irregularity between the columns within each expenditure category.

For the 'food' category, it is apparent that households in which at least one member is aged over 65, spend about 80% of the population average household spending level. As incomes rise, this figure is largely unchanged which indicates that this group increase spending on food at about the same rate as the population average. The family with children under 5, spend more than this initially and increase spending more rapidly as incomes rise, almost reaching the same level as the average for the population by the end of the simulation.

The older households spend slightly over 60% of the average on 'recreation' at the start of the simulation. The rate of increase is much faster than the population average and reaches over 120% of average spending after incomes were inflated. By contrast, the family with young children allocates less spending than the population average to this category as incomes rise. This leads to a decline in budget share for 'recreation' compared with the average household levels. In this way, the graph gives an indication of the relative priority of allocating the extra spending due to rising incomes.

5.4.3. Analysis by Income Quintile

Figure 27 shows changes in spending patterns by quintile as incomes rise, with quintile 1 being the lowest income group and quintile 5 being the highest. This is equivalised as before by dividing income and spending by the square root of household size. Results are presented for all 12 of the EFS high-level categories, as a percentage of the total spending among these categories.

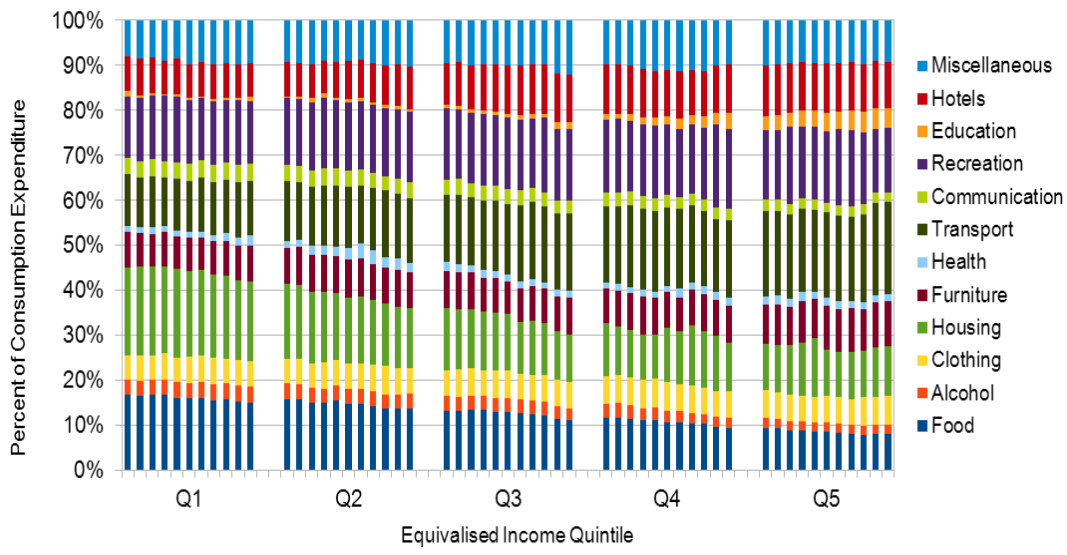


Figure 27: Share of Expenditure by Quintile

In Table 28 above (section 5.4.1), 'education' was identified as experiencing one of the largest increases in budget share. Figure 27 shows that the increase is moderate in the lower three quintiles. Much of the change appears in quintile 4 and for quintile 5, 'education' forms a significant budget item. Spending on 'transport' increases throughout to almost double its share between quintile 1 and 5.

The budget share for 'housing' (which includes rent and energy bills but not mortgages) declines as incomes rise, with much of this taking place in the lowest three quintiles. The proportion spent on 'food' declines as would be expected from Engel's Law.

5.5 Discussion

This example shows how a random assignment scheme can be applied to model a set

of several goods at a disaggregated level. The matching scheme used here selects a donor case and from there, any variables from the donor can be copied to the receiver. As a result, there is no restriction on the number of goods that can be modelled. In principle, every item in the EFS could be represented and the complexity of the model would rise linearly with the number of goods. The dimensionality problem, as described in Chapter 3 is avoided. Also, since the individual households are retained, the distribution of variables is preserved and output can be produced at the micro-level.

This does not mean the method is without pitfalls. As was mentioned above, selecting cases on a parameter such as increasing income means that all the variables from the case will be available for copying and some of these will be correlated with income. If the change in income really causes the change in another variable, this can be used to determine the effect of the former on the latter. However, if the change in income does not cause the change, such as when income does not cause household size to increase, this will lead to an inaccurate estimation of the effect because the simulated households change in a way that real households do not. This can be minimised by including more variables in the matching criteria to limit the household's response to changing conditions. A sensitivity analysis can be done by adding matching criteria, one by one, until the change in output reduces to an acceptable level. A regression of the matching variables on the output variables will indicate how much of the variation is explained by the model and so give an indication of its reliability. The problem of model specification is not unique to random assignment and all modelling methods

can be done well or badly whatever approach is taken.

Although random assignment preserves the individual cases, the model described above does not model trajectories. It is based on a cross-sectional dataset and assumes that households will respond to a change in circumstances by adjusting their spending patterns to become more like those that have already experienced the new conditions. This may have some plausibility but it is also reasonable to suppose that households would try to retain as much of their original behaviour as possible. It would be feasible to model this by using a longitudinal dataset and copying from donors that have experienced a similar change over time. Unfortunately, the EFS is a repeated cross-sectional survey and there are currently no large-scale longitudinal panel surveys, in the UK at least, that monitor detailed household spending categories over time.

5.6 Conclusion

Modelling household expenditure is usually done by parametric demand systems but there are some difficulties with this approach. These include the problem of modelling at a highly disaggregated level and limits on the number of goods that can be represented. The example model described in this chapter shows that a random assignment scheme provides one way to avoid these limitations. The chapter set out to illustrate a practical application of random assignment and the results point to areas for further research such as the possibility of matching using longitudinal data.

In order to isolate the effect of income on household expenditure, the demographic make up of the population was kept constant. The next chapter adds a random assignment component to the demographic model developed in Chapter 4, to estimate the effect of demographic change on household expenditure patterns in the context of what has been a major area for microsimulation modelling which is population ageing.

Chapter 6: Demographic Change and Expenditure

6.1 Introduction

The aims of this research, which were outlined in the opening chapter, included the evaluation of random assignment in terms of: feasibility (whether and how conveniently it can be done), validity (whether it produces the expected results) and flexibility (the range of situations to which it can be applied). The feasibility and validity of random assignment were tested in the previous chapter by implementing a simple model to project the effect of changes in household income on expenditure patterns. This chapter focuses on evaluating the flexibility of random assignment by applying it to a substantive problem in microsimulation modelling which is to investigate the implications of population ageing for household spending patterns. As such, it takes the form of a relatively self-contained case study in the application of a random assignment scheme using NetLogo.

In addition to its role in testing an application of random assignment, this chapter also contributes to the debate on population ageing by assessing its effect on aggregate household expenditure. It begins by reviewing how spending behaviour has been studied using both microsimulation and demand system approaches. The next section describes how the dynamic microsimulation model developed in Chapter 4 is used to provide demographic projections which are then built upon by a random assignment method similar to the one described in Chapter 5, to assign expenditure patterns to households as they change their demographic composition. This exposition covers

model specification and a simple alignment procedure to ensure that the population changes in a way that is consistent with ONS assumptions of changing mortality and fertility rates. The program algorithm is described next with the emphasis on the copying and matching procedure.

As the population ages, incomes are more predominantly copied from households where the members are retired. The effect of this, on average and aggregate household income, is modelled by including household income in the copied variables and the results for this parameter shown first. Following this, expenditure projections are presented for 'total expenditure' and the 12 high-level COICOP categories, as a percentage of aggregate UK expenditure in the base year of 2006. The next section describes the process of 'impact validation', which was promised in Chapter 4, to verify that the effect of demographic change on household expenditure is being modelled correctly. This consists of checking the arithmetic of the program, comparing the results against previous work where possible and finally assessing the effect of changes in model assumptions to determine its sensitivity to errors in the estimation of transition probabilities. Next, further results are presented for selected goods, disaggregated by age band and household type. Then some alternative scenarios are presented which show the effect of economic prosperity and austerity in comparison to the effect of demographic change. Some illustrative results of highly disaggregated categories are presented next and the chapter concludes with a discussion of the implications of population ageing on household expenditure patterns and the possible effect on the wider UK economy. An evaluation of the application of

random assignment and NetLogo in this context are left to the final chapter.

6.2 Demographic Change and the Ageing Population

It was noted, in Chapter 2, that one of the main reasons for developing dynamic microsimulation models has been concern over the effects of structural population ageing where increasing life expectancy leads to a rise in the proportion of pensioners compared with those of working age. According to Brown et al. (2011), the debate on population ageing can be divided into two sides. One is the 'crisis' perspective which sees the ageing population as being unaffordable and a threat to the stability of welfare systems. The other is the 'manageability' perspective which argues that by taking into account a range of other factors, such as the compression of morbidity and increased industrial productivity (Mullan, 2000), the ageing population presents less of a problem that it appears by considering its costs in isolation. As discussed in section 2.4, this issue has been studied with a range of models to determine the magnitude of the additional costs for healthcare and pensions that an ageing population is expected to bring.

While it has been the subject of a number of studies in economics, as described in Chapter 2, the effect of a change in the age distribution of the population on household spending patterns may also have a significant economic impact but seems to have received rather less attention from the microsimulation community. Of the few microsimulation models that contain an expenditure component, EUROMOD and SPIT (Section 2.5) are both static models and so do not incorporate the effects of

demographic change. The dynamic microsimulation model developed by Ando and Altimari (2004), models total consumption but this is not disaggregated to show which type of goods might be affected to a greater or lesser extent.

6.3 Combined Demographic/Expenditure Model

In common with the income model of the previous chapter, the combined demographic/expenditure model takes the 2006 EFS as its base population. This dataset, is weighted to form a representative sample of the UK and has a large number of expenditure categories along with some demographic information about each household. The initial population is advanced over time using the annual transition probabilities that were obtained and used in Tyche, described in Chapter 4. At the end of each simulated year, it is then possible to model the new expenditure pattern of the population using a random assignment scheme. The model operates on the simple principle that whenever a household experiences a change in its demographic composition, the income and expenditure pattern is copied from another household that had the same or similar composition in the previous year. This is based on the underlying assumption that whatever adjustment is to be made by the changing household will already have been made by the donor household. Alternatively, whatever behavioural rules were applied by the donor household to allocate expenditure to items, will be implied in the new expenditure pattern of the recipient household.

6.3.1 Expenditure Model Specification

The random assignment component of this model is derived from the income model described in Chapter 5. However, instead of copying from a donor case that has a different income, the donor is defined to be a household where the oldest person is one year older than in the current household. The age of the oldest person in the household was found to be more convenient to define in program code for newly created households than a notional 'head of the household' or the 'household reference person' (HRP) as used by the EFS. In that survey, the HRP is defined as the person who owns the accommodation, is responsible for paying the rent or who is in some way legally responsible or entitled to the accommodation. While the HRP is identified in the base data set, the demographic modules create some new households and if this variable was to form an indicator of the age of the household, one of the occupants would have to be designated as the HRP. However, the individual-level data need to make an accurate selection of the HRP is not available and an inaccurate assignment of HRP status could lead to an artefact where the characteristics of the head of the household change over time. Inspection of the EFS showed that the variable a070p (age of the oldest person in the household) has the same value as p396p (age of household reference person) in 81% of cases and is within three years in 91% of cases. However, the priority is not to mimic the designation of the HRP exactly but to obtain a consistent measure of the age of the household. Further research that investigates the effect of different proxies for age of the household might be informative but age of the oldest person appears to be adequate in the current context.

The constraining variables of 'household size' and 'household type' are also included as matching criteria, as was the case in the income model of Chapter 5. The copying mechanism allows the model to implicitly take account of the effect of changing incomes as people age. As expenditure patterns are copied from older households, the budget set is drawn from a population that has the income distribution of the older group. This means that as the proportion of pensioners in the population increases, the expenditure patterns are more frequently selected from a household in which the oldest person was a pensioner in the base data set. It is therefore possible to copy the income from the donor case and then aggregating incomes across the population indicates how this changes, assuming constant base year prices and no change in pension policy. This is an example of how the random assignment scheme models unobserved heterogeneity.

6.3.2 Alignment

The transition probabilities for births and deaths are adjusted so that their rates approximate ONS assumptions. This was done using the method described in Chapter 4. There are also sliders on the user interface that facilitate changing birth and death rates if desired, to accommodate a range of demographic scenarios. These are used later to test the sensitivity of the results to these assumptions.

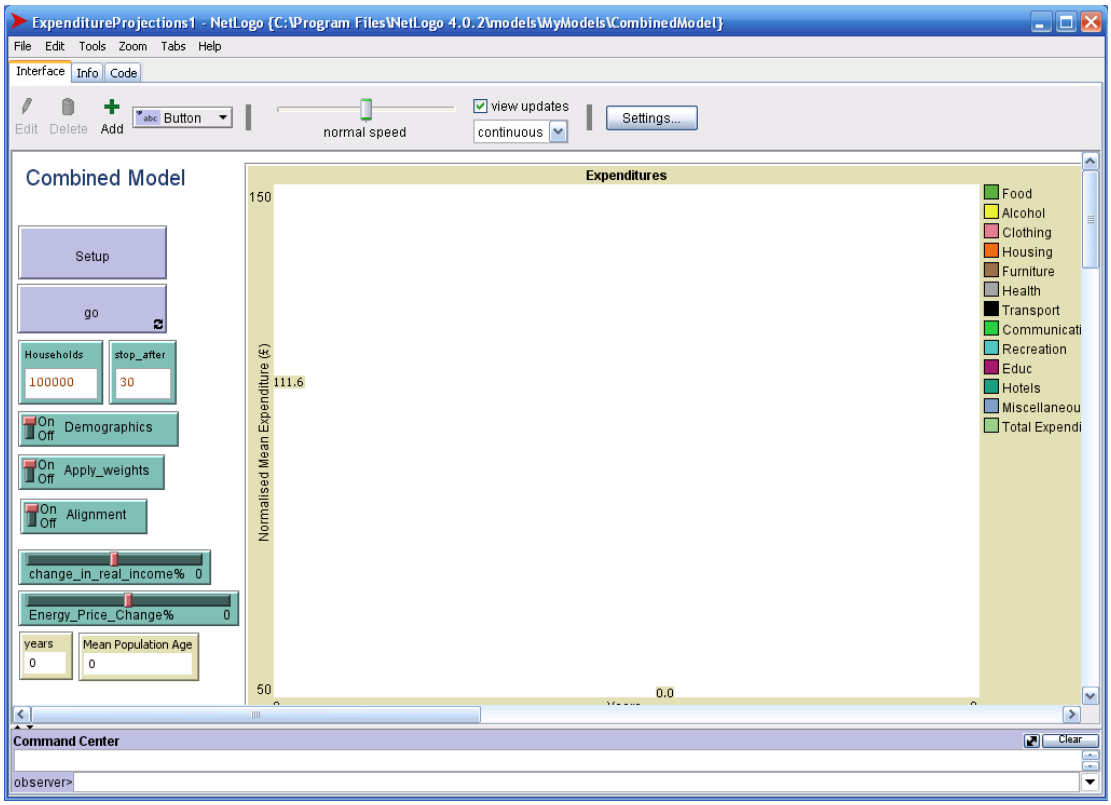


Figure 28: Combined Model User Interface

6.3.3 Running the Model

When the program is initiated, the user is presented with the screen shown in Figure 28. From here it is possible to set the number of households to load, the number of years the simulation is to run for and whether alignment is to be applied. When this is done, the 'setup' button initiates loading of the base data files of households and individuals. After loading the appropriate number of cases, the 'go' button begins the dynamic simulation. Each simulated year, commences by running the demographic modules of mortality, fertility, leaving home, partnership formation and dissolution. This results in a new cross-sectional population, advanced by one year, which is then processed through the expenditure module. For each household, in random order, the

income and expenditure pattern of a household where the oldest occupant is one year older than that of the current household and also had the same number of occupants and household type is stored ready to update the current household. If there are no households that match on all three criteria, matching takes place on age and household size. If this also fails, a household where the oldest occupant is one year older than in the current household is used as the donor. When using the full EFS, the number of failures to match on all three criteria is negligible. Once all households have been processed, the stored expenditure patterns are copied into the household variables. When all updates are complete, aggregate totals for variables of interest can be calculated and plotted on graphs, displayed on the screen or written to a file for further analysis. This process is summarised in pseudocode as shown below.

```
set model parameters
load cross sectional household data file
load individual level data file
assign individuals to households using the household identification number (HID)
for each year
   for each household
      update demographic characteristics for 1 year
   for each household
      locate a household where the oldest person is 1 year older
                                    than the oldest person in the current household
                                    which has the same number of occupants
                                    and household type
      store its expenditure pattern
   for each household
```

update expenditure pattern

calculate new aggregate expenditures for goods of interest

6.4 Results

The combined model forms a coherent micro-level framework for the analysis of household expenditure patterns. As such, it permits the modelling of a virtually unlimited range of demographic scenarios by changing the transition probabilities for births, deaths, leaving home, partnership formation and dissolution, to determine how they would affect household spending. It is also possible to vary economic parameters such as household income, as will be done later in this chapter, and the results can be produced at any level of disaggregation. Some illustrative results are given below. The results presented in this section provide an example of the output that can be obtained for the scenario of an ageing population based on ONS assumptions of fertility and mortality rates. Some initial results of how demographic change is projected to affect the age and household size of the UK population were provided in Chapter 4. In this section, the random assignment component is run after the demographic element and used to project aggregate spending for 'total expenditure' and the twelve high-level COICOP categories. In order to isolate the effect of population ageing, all factors apart from demography are assumed to remain the same. However, as more people move into retirement, this has an endogenous effect on both average and aggregate household incomes. The model also allows implicitly for the effect of households drawing on savings, by the same mechanism and this unobserved heterogeneity is embedded in the results for expenditure presented here.

6.4.1 Changes in Household Income

In Chapter 4, it was projected that natural change in the UK population would lead to a 12% increase in the number of households. All other things being equal, this would be expected to lead to a proportionate increase in aggregate spending. However, it was also projected that household occupancy is falling and the proportion of retired people will be increasing. The reducing household size and lower income that people often experience in retirement would tend to have a restraining effect on both income and spending. It is therefore possible to estimate the endogenous changes in household income due to population ageing. Figure 29 shows three series which are the result of three simulations. The difference between each series is due to stochastic variation and allows an estimate to be made of the level of uncertainty this introduces into the model results.

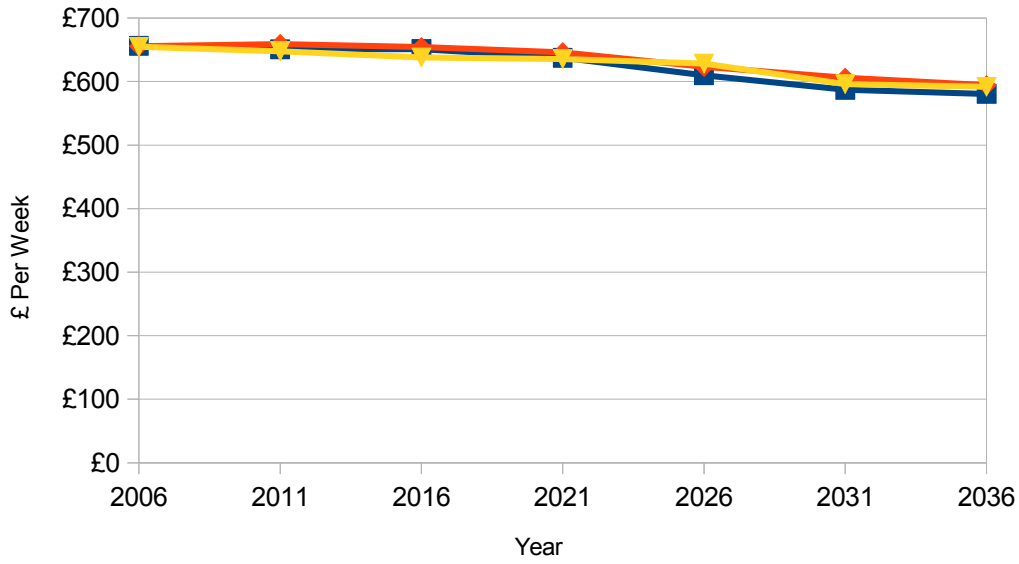


Figure 29: Endogenous Change in Average Household Income: 2006 - 2036

Figure 29 indicates that there is a moderate fall in average household incomes which would be expected as a result of the increasing proportion of retired people.

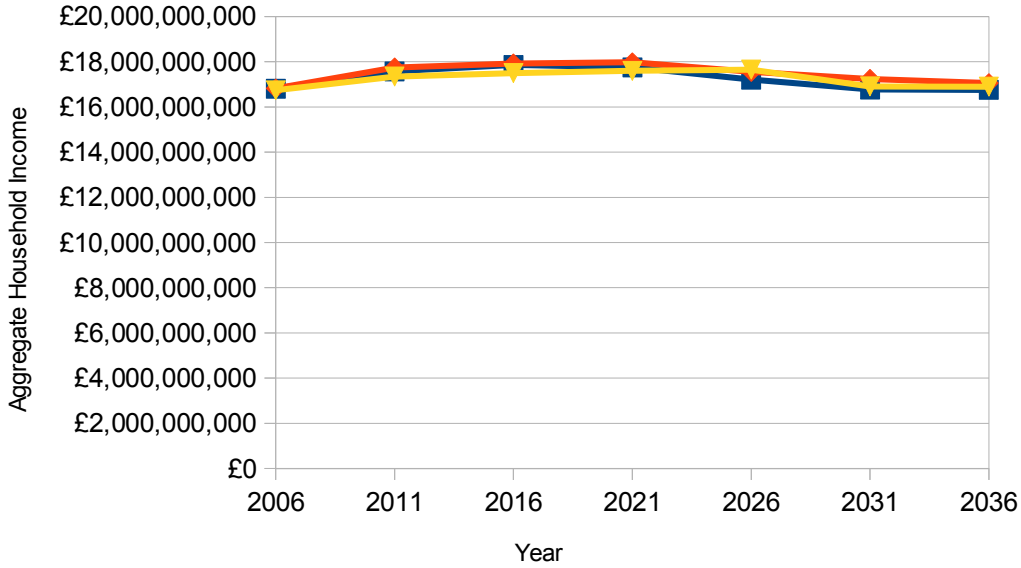


Figure 30: Endogenous Change in Aggregate UK Total Household Income

Figure 30 shows a small rise in aggregate household income due to the increasing population. This is reversed after 2021 when falling average incomes offset the growth due to the greater number of households.

6.4.2 Spending Patterns Over the Life-course

Luhrmann (2008) notes that, since household-level consumption patterns change over the life cycle, aggregate demand is likely to react to a shift in the population age structure. Figure 31 shows how expenditure patterns in the UK change with the age of the oldest household occupant using data from the 2006 EFS ('total expenditure' is divided by 10 to fit on the scale).

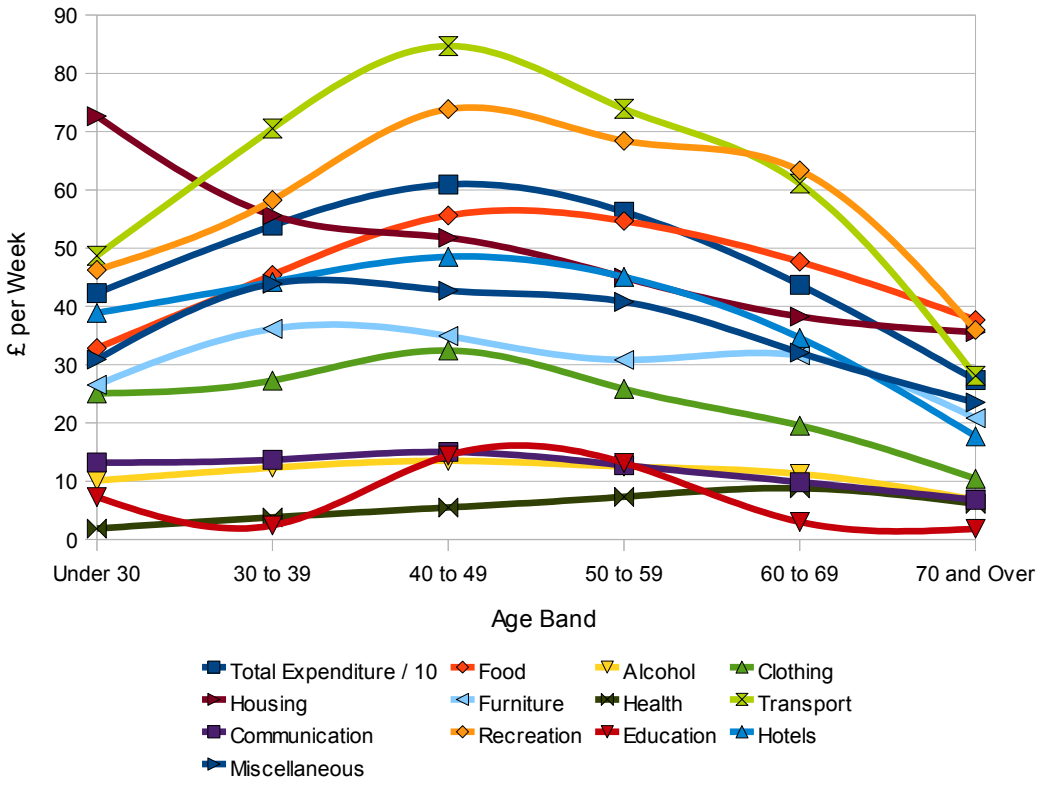


Figure 31: Spending Patterns for Different Age Groups in 2006

Expenditure for most goods begins at a relatively low level and then increases to a peak in the '40 to 49' age band. It remains high for the '50 to 59' age group before declining in the oldest households. Some goods depart from the usual 'inverted U' shape. 'Housing' is at its highest for the 'under 30' group and declines for all age bands after that. 'Health' peaks in the '60 to 69' age group. 'Education' has two peaks, one for the 'under 30's who may be taking courses to develop skills and one in the '40 to 49' age band who may presumably be paying for their children's education.

6.4.3 Projection of Household Expenditure

The combined model was run incorporating 'total expenditure' and all 12 high-level

COICOP categories recorded in the EFS. It is common practice in economics to present spending patterns in terms of average budget share per household (Paris and Houthakker, 1955) (Barigozzi et al., 2009). In effect, this represents the result of dividing total spending for all households by the number of households and then dividing by total income or expenditure to express the results as a share of an average household budget. The effect of this is to cancel out variations in population size and income over time because, usually, the interest is in how a typical household will allocate each unit of expenditure. However, in this case, since total expenditure varies over the life cycle, as was shown in Figure 31, aggregate expenditure can be expected to vary in response to a shift in the age distribution of the population (Luhrmann, 2008). Moreover, population size is changing over the period due to changes in the birth rate and life expectancy. Cancelling this out by dividing by population size would negate much of the purpose of demographic modelling. For these reasons and following Takeuchi's (2005) approach, the results of the simulations are plotted for aggregate expenditure where each good is set to 100 in the base year of 2006 and subsequent years are compared to this. As a result, all prices are nominally fixed at their 2006 values.

The stochastic nature of random assignment means that the outcome of each simulation is slightly different each time the model is run. The results presented below represent the average of ten simulations. 95% confidence intervals were calculated to form a measure of the amount of variation between each run. As the process generating these results is a type of discrete Markov process (Andreassen,

1993), it can be expected that the results are drawn from a normal distribution. This was tested for 'Total Expenditure' in the years 2011 to 2036. A Shapiro-Wilk test for normality was conducted using SPSS with results shown in Table 29.

Total Expenditure for Year	Shapiro-Wilk Statistic	df	Significance
2011	.958	10	.768
2016	.960	10	.786
2021	.900	10	.220
2026	.954	10	.721
3031	.954	10	.712
2036	.850	10	.061

Table 29: Analysis of Distribution of Total Expenditure over 10 Simulations

In all cases, the level of significance is above the threshold of 0.05 and so does not indicate a rejection of the null hypothesis that the data is consistent with a normal distribution. Aggregate expenditures are reported on the graphs below.

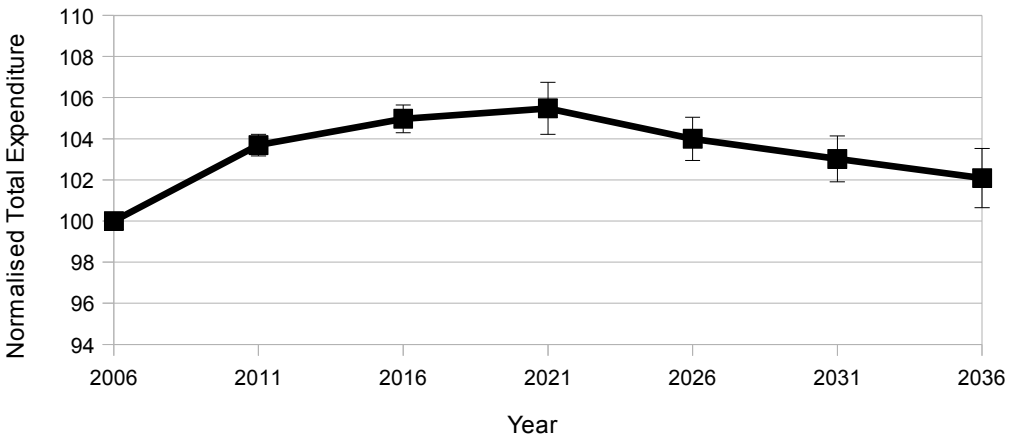


Figure 32: Total Expenditure

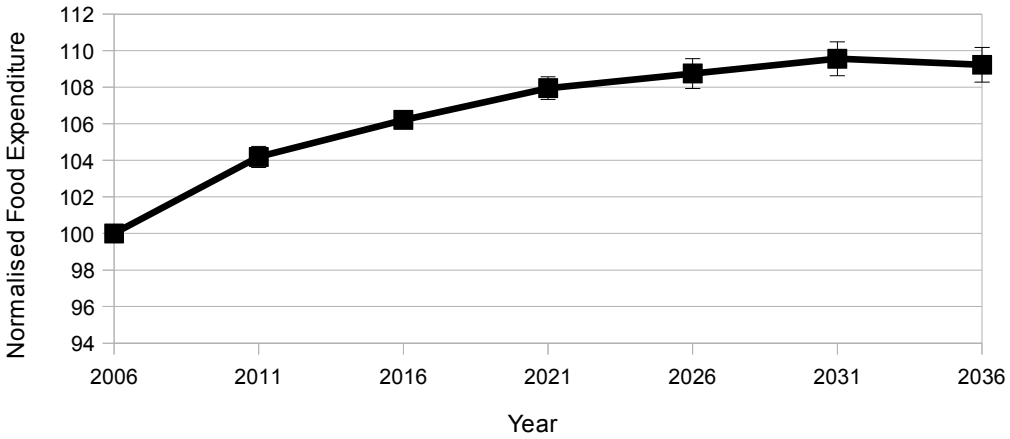


Figure 33: Food (food & non-alcoholic drinks)

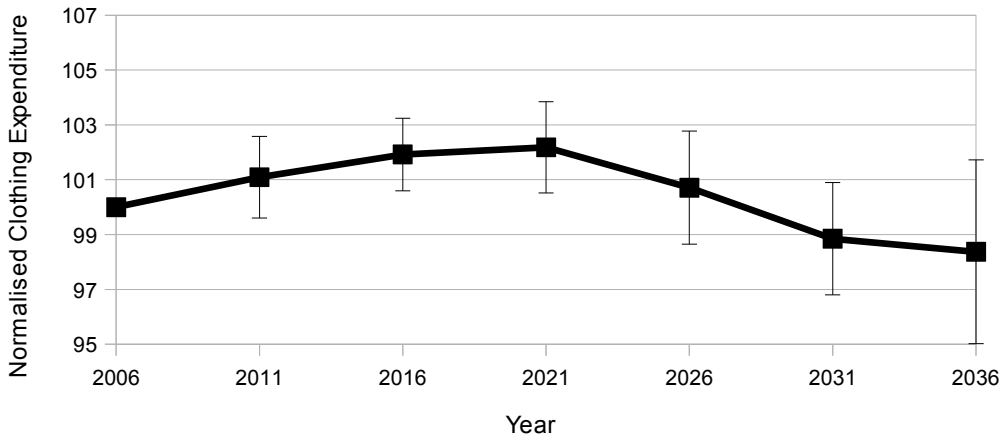


Figure 35: Clothing (clothing & footwear)

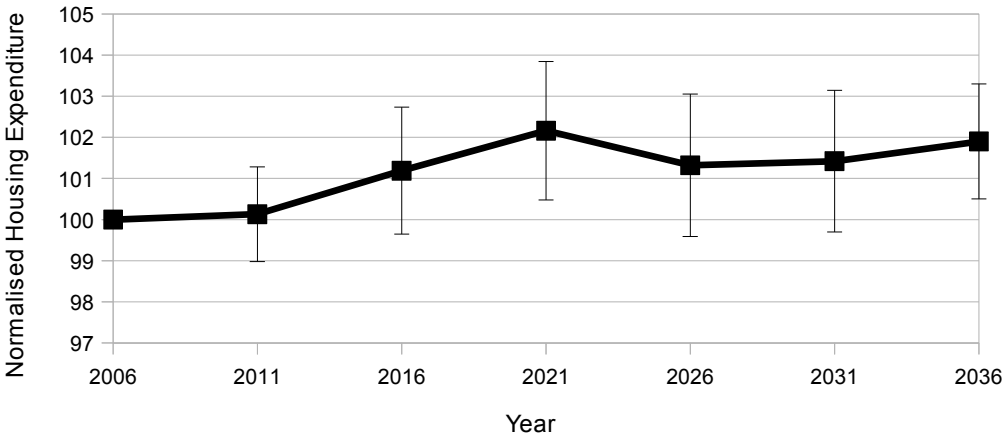


Figure 36: Housing (housing, fuel & power)

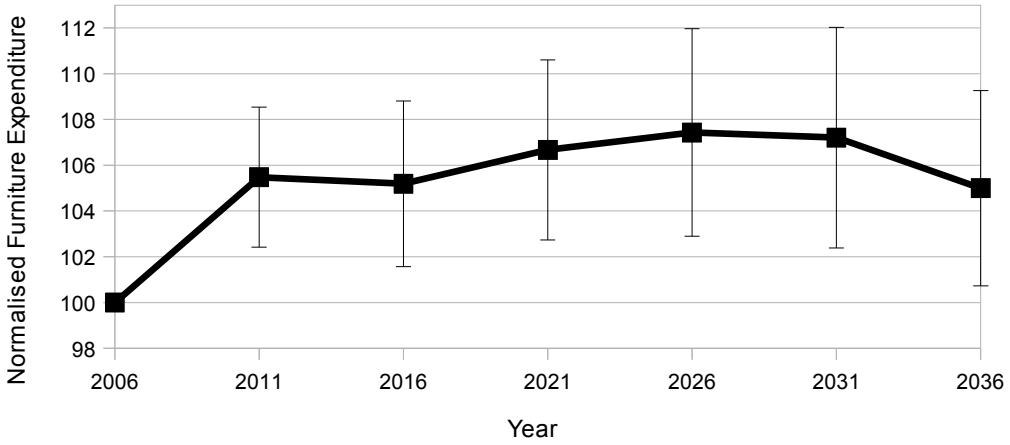


Figure 37: Furniture (household goods & services)

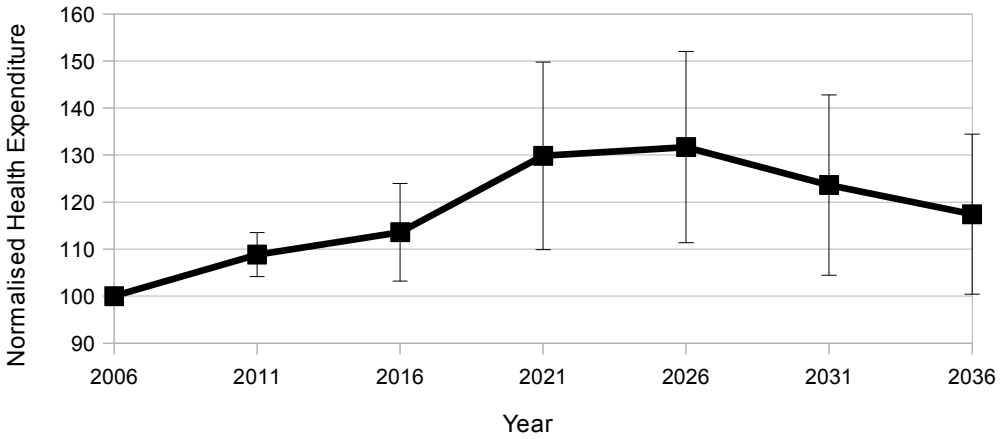


Figure 38: Health

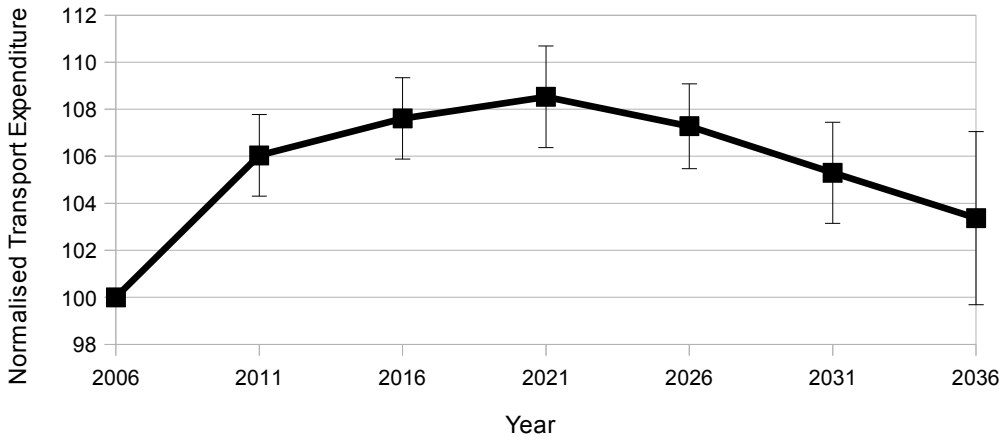


Figure 39: Transport

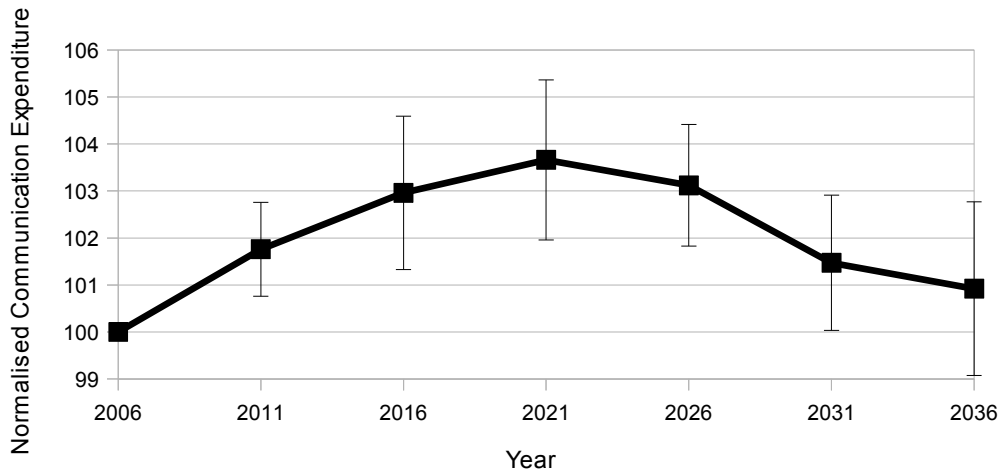


Figure 40: Communication

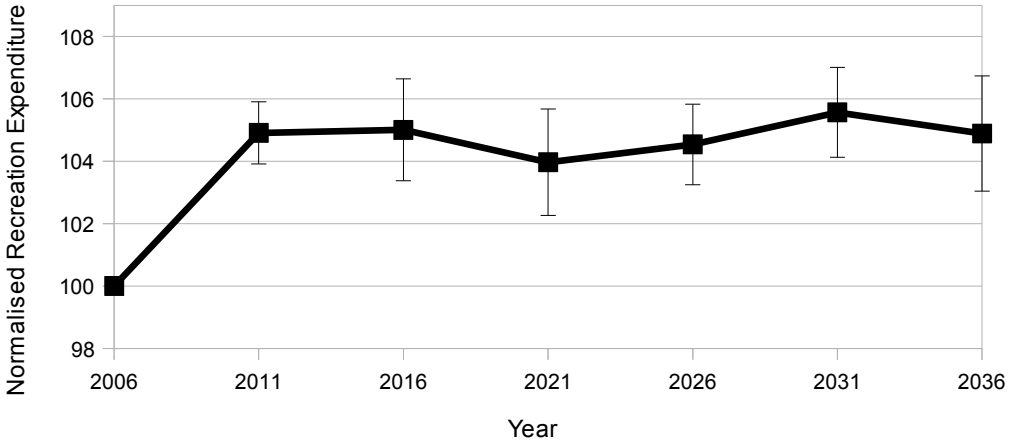


Figure 41: Recreation (recreation & culture)

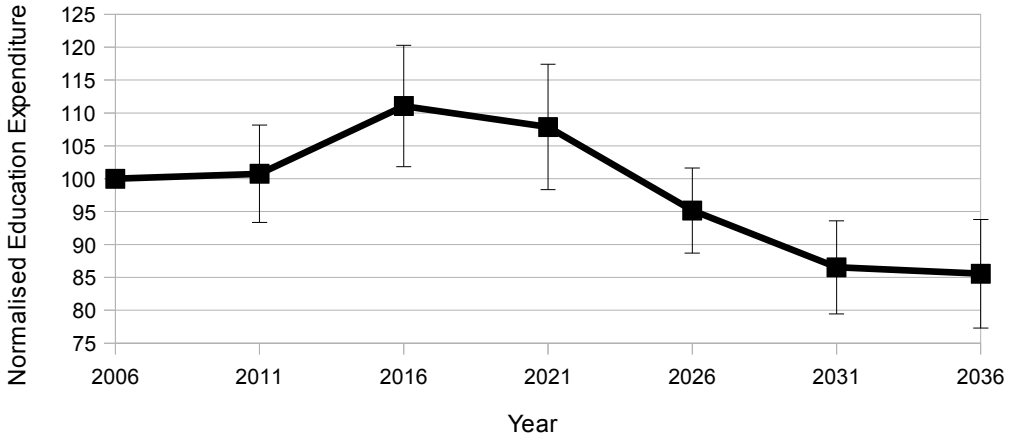


Figure 42: Education

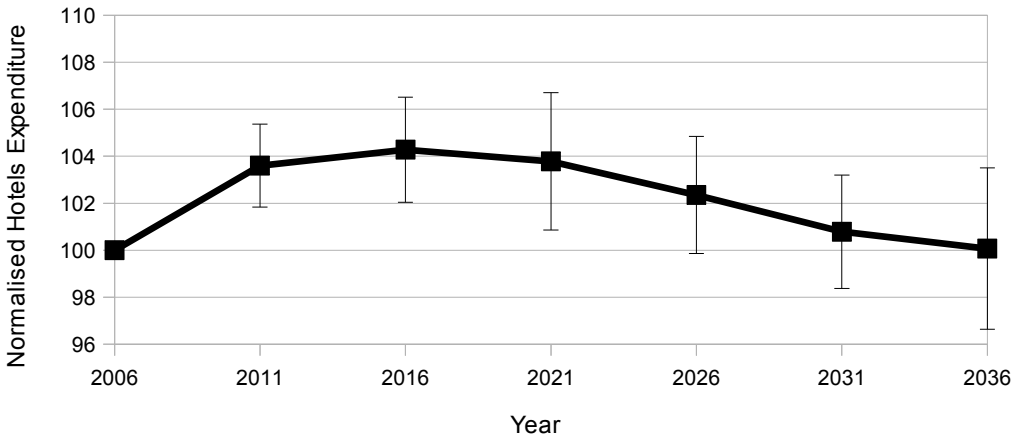


Figure 43: Hotels (restaurants & hotels)

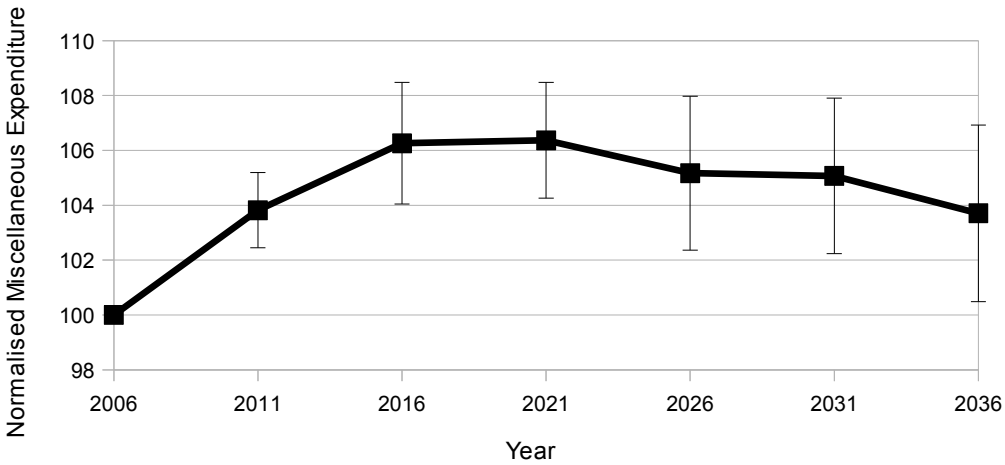


Figure 44: Miscellaneous (miscellaneous goods & services)

The results are summarised in Table 30 which, following 'total expenditure', is sorted into descending order of aggregate expenditure in 2036. Spending categories near the top of the list are favoured by population ageing. Those near the bottom experience more moderate growth and some decline by 2036.

	2006	2011	2016	2021	2026	2031	2036
Total Expenditure	100	103.7	105	105.5	104	103	102.1
Health	100	108.8	113.6	129.8	131.7	123.6	117.4
Food	100	104.2	106.2	108	108.8	110	109.2
Alcohol	100	104	105.1	106.2	105.7	106	105
Furniture	100	105.5	105.2	106.7	107.4	107.2	105
Recreation	100	104.9	105	104	104.5	105.6	104.9
Miscellaneous	100	103.8	106.3	106.4	105.2	105.1	103.7
Transport	100	106	107.6	108.5	107.3	105.3	103.4
Housing	100	100.1	101.2	102.2	101.3	101.4	101.9
Communication	100	101.8	103	103.7	103.1	101.5	100.9
Hotels	100	103.6	104.3	103.8	102.4	100.8	100.1
Clothing	100	101.1	101.9	102.2	100.7	98.9	98.4
Education	100	100.8	111	107.9	95.1	86.5	85.6

Table 30: Normalised Aggregate Expenditure

Table 31 provides an analysis of changes in percentage share of total expenditure arranged with those with the largest increase at the top.

	2006	2011	2016	2021	2026	2031	2036
Recreation	12.5	12.73	12.91	13.01	13.29	13.47	13.33
Food	9.89	9.89	10	10.1	10.27	10.45	10.59
Housing	9.81	9.64	9.72	9.86	9.91	10.12	10.27
Transport	12.85	12.85	13.41	13.37	13.31	13.06	13.02
Miscellaneous	7.56	7.56	7.51	7.53	7.57	7.76	7.67
Alcohol	2.37	2.37	2.36	2.39	2.39	2.43	4.46
Health	1.3	1.3	1.25	1.36	1.41	1.37	1.38
Communication	2.45	2.45	2.38	2.39	2.41	2.41	2.42
Furniture	6.41	6.41	6.3	6.26	6.27	6.33	6.34
Hotels	7.85	7.85	7.88	7.86	7.89	7.77	7.76
Clothing	4.83	4.83	4.7	4.64	4.56	4.6	4.6
Education	1.44	1.41	1.43	1.35	1.26	1.11	1.17
Other	20.74	20.74	20.16	19.86	19.48	19.12	19

Table 31: Budget Shares (%)

6.4.4 Impact Validation

As described in Chapter 4, the final stage of testing APPSIM was 'impact validation'. The purpose of this was to develop confidence that the model provides an accurate simulation of the effect of a policy change. Part of this was done by verifying aggregate totals during the simulation to check the model's internal consistency. Further checks were made by comparing the output with results from other economic models. The approach taken to test the expenditure system in this chapter is also designed to develop confidence in the results and to show that they give an accurate reflection of the effect of changing demographics on household spending patterns. This is done in three stages. The first is to check that the file input, copying mechanism and the calculation of aggregate totals is being done correctly. Next, the outputs of the model are checked against some of the results from previous research. Finally, some alternative demographic scenarios are run to assess the sensitivity of the model to the assumptions of birth and death rates.

6.4.4.1 Totals Check

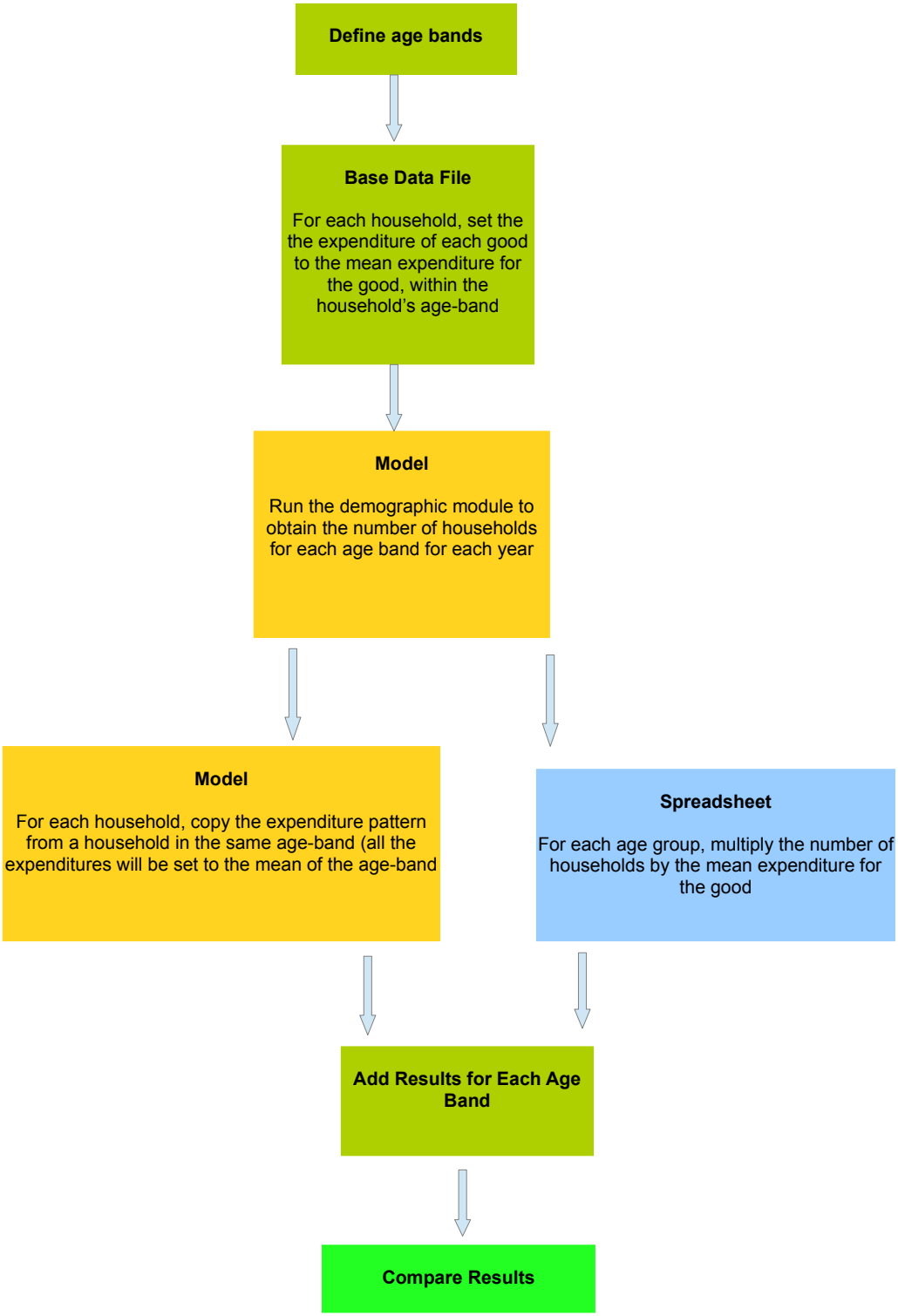
Program testing in a model based on random assignment is complicated by the stochastic nature of the process. One way to check the internal consistency of the program is to remove the random element and then compare the totals for aggregated groups against test data obtained by using another method; in this case, a spreadsheet calculation.

The demographic model allows the number of households to be projected over time,

by ten year age group, as was shown in Figure 17 in Section 4.8 above. The mean expenditure for selected items is also known for the base year of 2006, as was shown in Figure 31. With this information, it is relatively straightforward to use a spreadsheet to obtain the total expenditure for each item by multiplying the mean expenditure by the number of households projected for each year group. Adding up the totals for each year will then give the aggregate expenditure for each good.

These results can be reproduced in the NetLogo model by assigning each household to an age band and setting each expenditure to the mean value for a household in its age band. Then, whenever a household copies a new expenditure pattern, its spending on each good will be the average for its age band. Finally, aggregating across age bands will give the total expenditure for each good.

This process is illustrated below.



When this was done, the results were found to be identical except for an occasional

difference in the third decimal place due to differences in rounding conventions used by the spreadsheet, OpenOfficeCalc and NetLogo. This demonstrates that there are no errors, detectable using this method, that affect these particular results, in the file input, copying mechanism and the calculation of aggregate totals.

6.4.4.2 Comparison with Previous Results

While Section 2.5 above, noted several pieces of research into the effect of population ageing on household expenditure patterns, few of them are directly comparable with those produced here. The only one to report on UK expenditure was Luhrmann (2008) but the results did not take into account changing household size with age and were presented as budget shares, not aggregate expenditures. The same holds for Lefebvre's (2006) results for Belgium. Only Takeuchi (2005) gives results in terms of aggregate expenditures and these are summarised in Illustration 1.

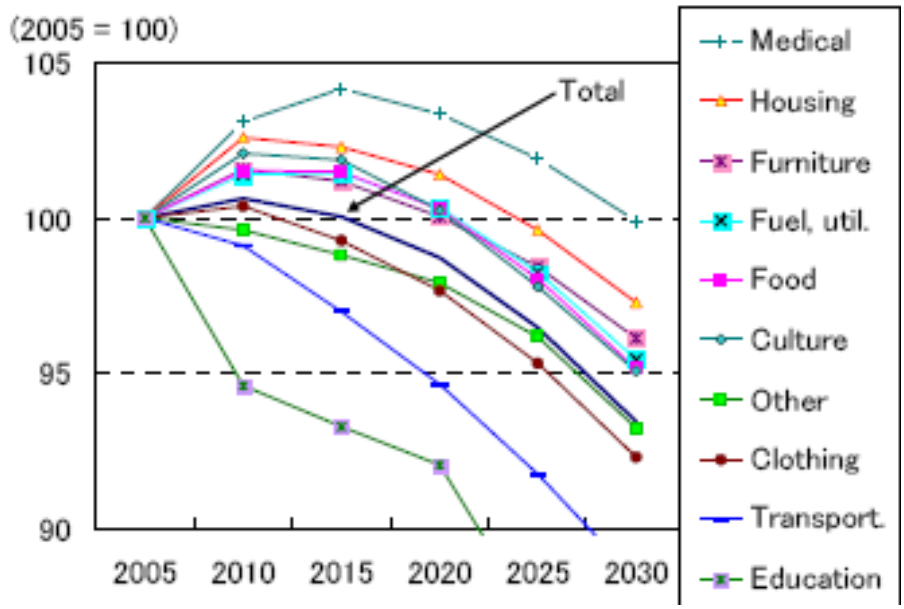


Illustration 1: Changes in Aggregate Household Spending in Japan (Takeuchi, 2005)

Although Japan is experiencing a similar ageing population phenomenon to that of the UK, there would be little reason to suppose that spending patterns in the two countries would correspond exactly. Nevertheless, Illustration 1 shows that for many goods, there is a rise in spending followed by a decrease, similar to what has been projected for the UK. This is not the case for all goods and five of the categories are below their 2005 baseline by 2030. Another similarity with the UK is that 'health' experiences the greatest increase and 'education' has the greatest fall. One difference is that while 'housing' experienced only a modest 2.2% increase at its peak in the UK projection, ranking it in the lower half of the 12 COICOP categories, 'housing' received the second highest (3%) increase after 'health' in Japan. Also, both Lefebvre and Luhrmann noted an increase in the budget share for housing or its corresponding category. One explanation for the more modest increases in 'housing' spending, found in the research described here, may be that these projections take into account changes in household occupancy within age bands over time. This could lead to the observation of smaller households with less expenditure within each age band.

6.4.4.3 Sensitivity Analysis

In population modelling, it is often useful to present results for a central or principal projection', based on what appear to be the most likely assumptions for the relevant parameters for example (ONS, 2013b). In addition to this, one or more alternative scenarios are produced to show the effect of small deviations from the base assumptions. This gives an indication of how sensitive the results are to the underlying assumptions or how robust the model is to changes in its parameters. In

Figure 45 below, results are presented for 'total expenditure' in the base scenario (which is aligned to ONS assumptions) and four alternative scenarios where the annual transition probability for each case is varied from its evaluated level by a fixed percentage. These are: 1) the probability of birth is increased by 50%, 2) the probability of birth is decreased by 50%, 3) the probability of death is increased by 50% and 4) the probability of death is decreased by 50%.

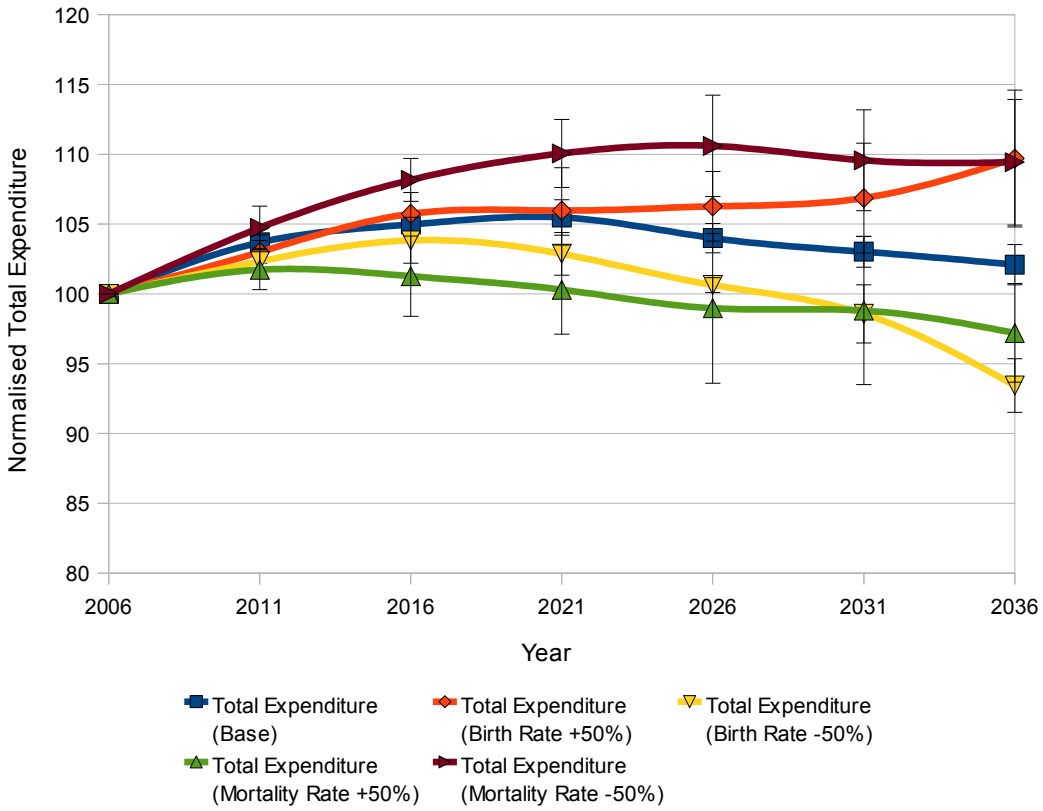


Figure 45: Effect of Alternative Transition Probabilities

It can be seen from Figure 45 that higher birth rates and lower death rates lead to higher aggregate expenditures, as would be expected due to an increased population. Likewise, lower birth rates and higher mortality rates decrease expenditure. The spread of results due to 50% variations in transition probabilities translates to less

than 10% in expenditure. This indicates that the results are relatively insensitive to assumptions of mortality and fertility rates. Part of this is due to the fact that only a small proportion of the population are born or die in a particular year so that variations in the probability of these events manifests slowly. Another factor is that births and deaths tend to take place at the ends of the age distribution when spending is lower than in middle age. This tends to attenuate the effect of changes in birth and death rates on aggregate spending.

6.5 Disaggregated Results

The micro-level datafile of individuals and households, which is available after every year of the simulation, allows the results to be presented in a wide variety of ways. The following sections provide some results in which the population is decomposed into selected groups of interest. The first is by ten-year age bands because, as the focus of this model is population ageing, it seems natural to study how spending changes within different age categories. Next spending is disaggregated by industrial sector, first by age band, then household type. This kind of output could be useful for a manufacturer or producer, to anticipate how the characteristics of consumers might change as a result of population ageing and then alter their marketing strategy or product range in response. Analysis of the results is left to the discussion in Section 6.8 below.

6.5.1 Spending by Age Band

Rather than present results for all 12 COICOP categories plus ‘total expenditure’,

which would become repetitive, detailed projections have only been produced for ‘housing’, ‘health’, ‘transport’ and ‘food’. ‘Health is of interest because it is an item that can be expected to increase with population ageing and is also unusual in that it peaks for the ‘60 to 69’ age band. ‘Housing’ is included because of the projected shift in household occupancy towards smaller households and spending in this category peaks in the youngest age group. ‘Food’ is a staple and would tend to increase with the number of individuals rather than households. ‘Transport’ is the largest spending category for the middle age groups but decreases quickly for the oldest so appears quite sensitive to age.

Figure 46 gives an overview so that all categories can be seen on the same scale. Within each category, annual aggregate expenditure is shown for the years 2006 to 2036 in five year intervals.

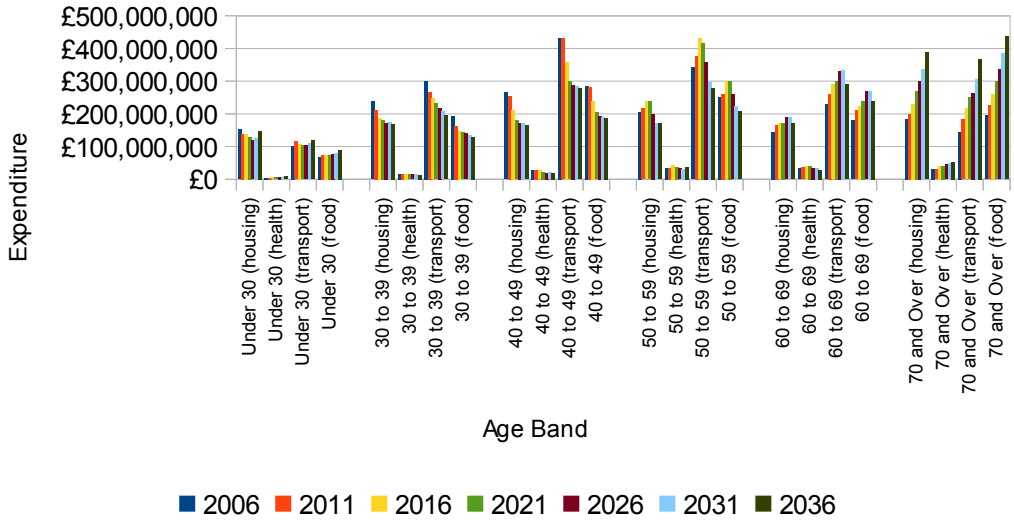


Figure 46: Expenditure by Age Band (housing, health, transport, food)

The next four graphs show how spending for ‘housing’, ‘health’, ‘transport’ and ‘food’ vary by age group between 2006 and 2036. The results are the average of three simulations showing 95% confidence intervals. As this microsimulation is a discrete Markov process, the results are normally distributed and deriving confidence intervals from a small sample is valid.

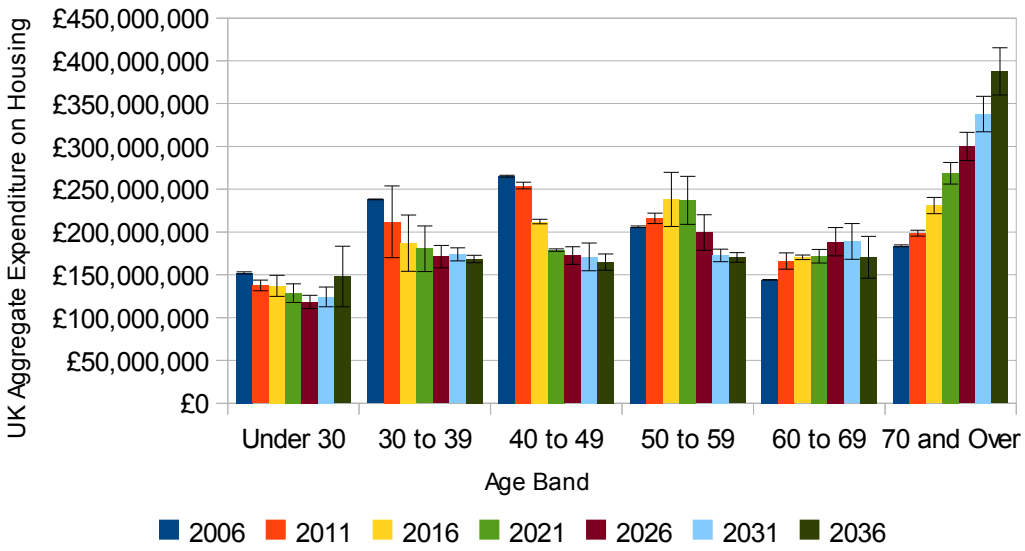


Figure 47: Aggregate Spending on Housing by Age Band

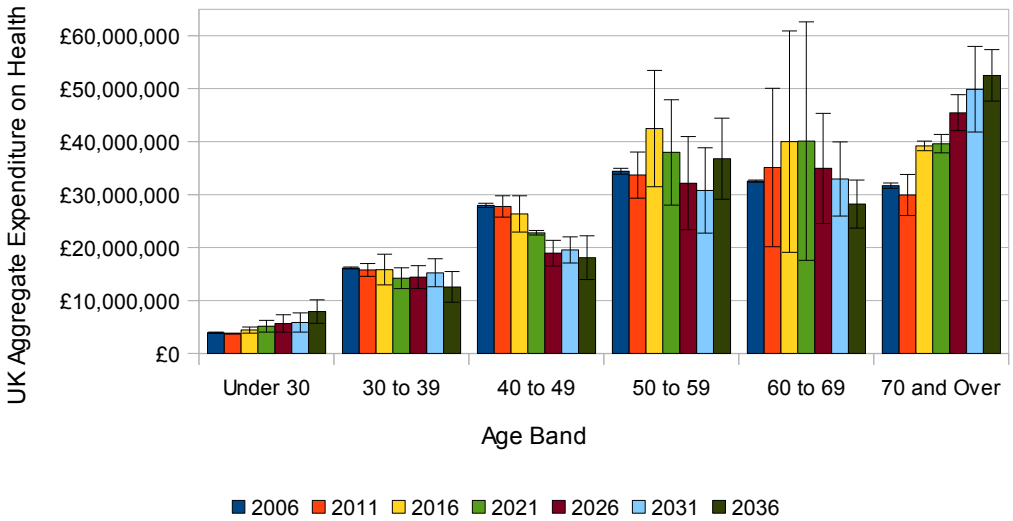


Figure 48: Aggregate Spending on Health by Age Band

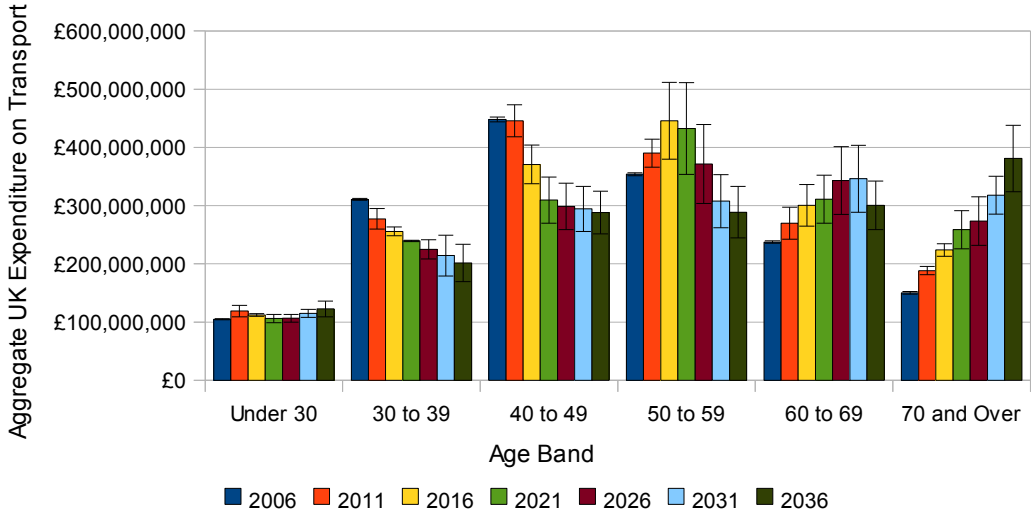


Figure 49: Aggregate Spending on Transport by Age Band

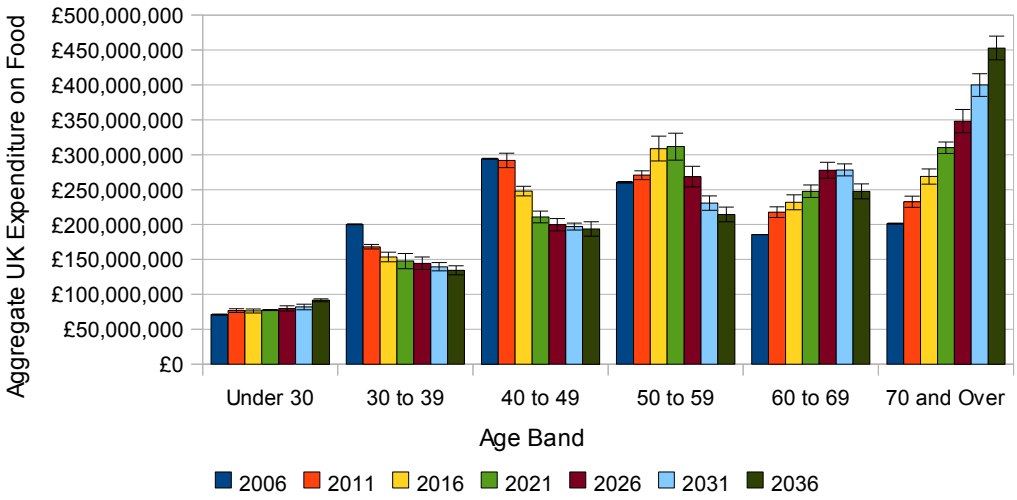


Figure 50: Aggregate Spending on Food by Age Band

6.6 Spending by Industrial Sector

While the data provided by the EFS is given at the individual and household level, aggregate spending within each COICOP category can also be used as an indicator of the level of demand for goods within various industrial sectors. The next series of

results are based on the same data that was presented above except spending by each age group is represented as a percentage of total projected expenditure for that category. This can be used to infer how the characteristics of consumers for each sector will change over time given the assumptions used in the model. In this section, results are presented for all 12 COICOP categories.

6.6.1 Spending by Age Band

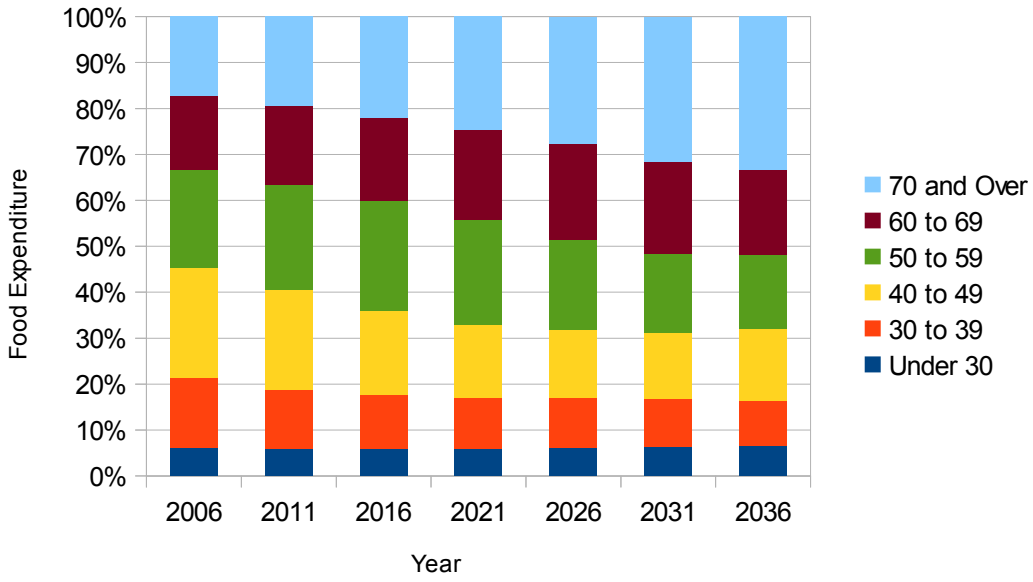


Figure 51: Share of Food Spending by Age Band

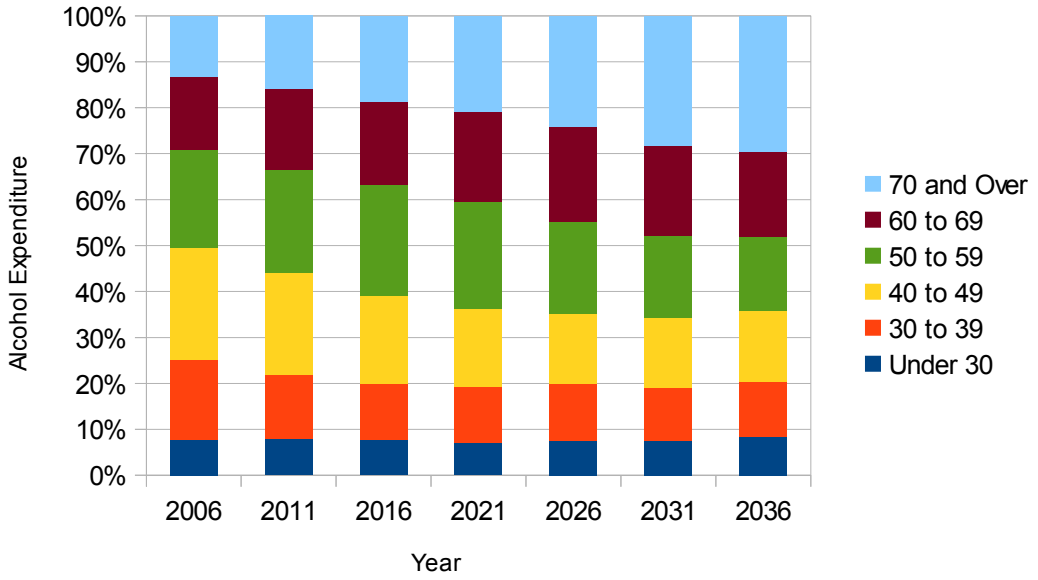


Figure 52: Share of Alcohol Spending by Age Band

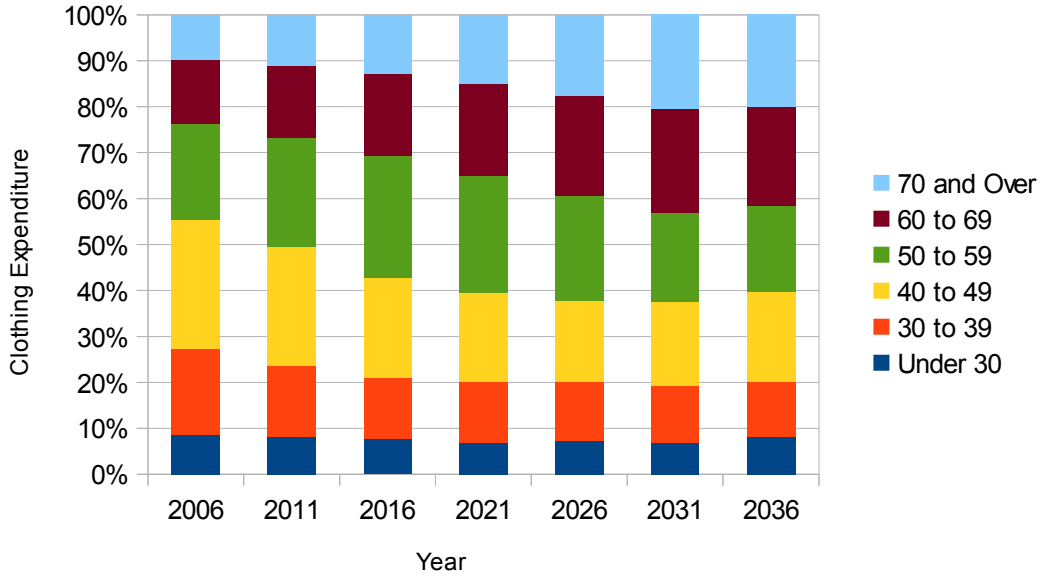


Figure 53: Share of Clothing Spending by Age Band

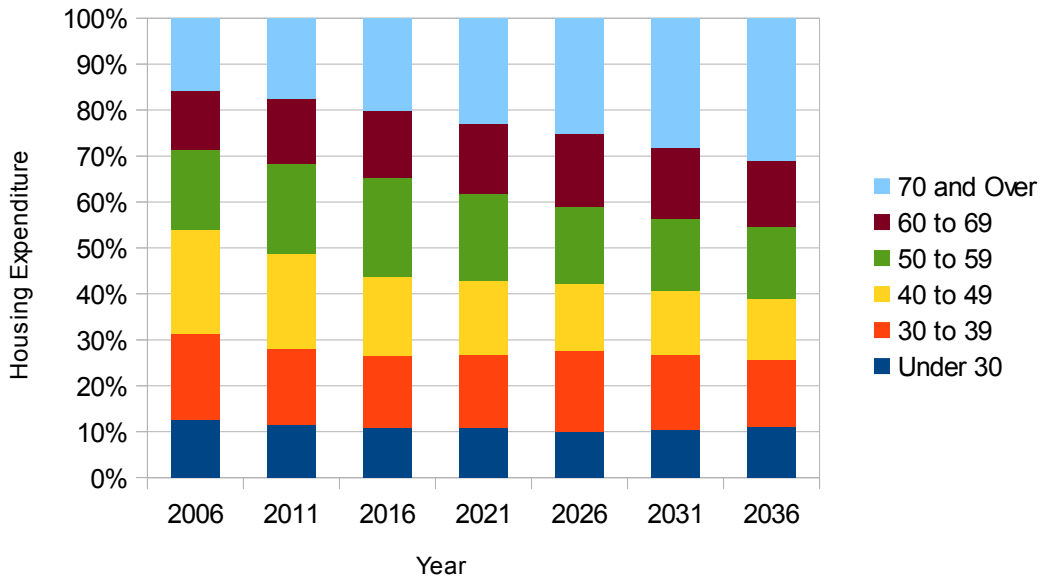


Figure 54: Share of Housing Spending by Age Band

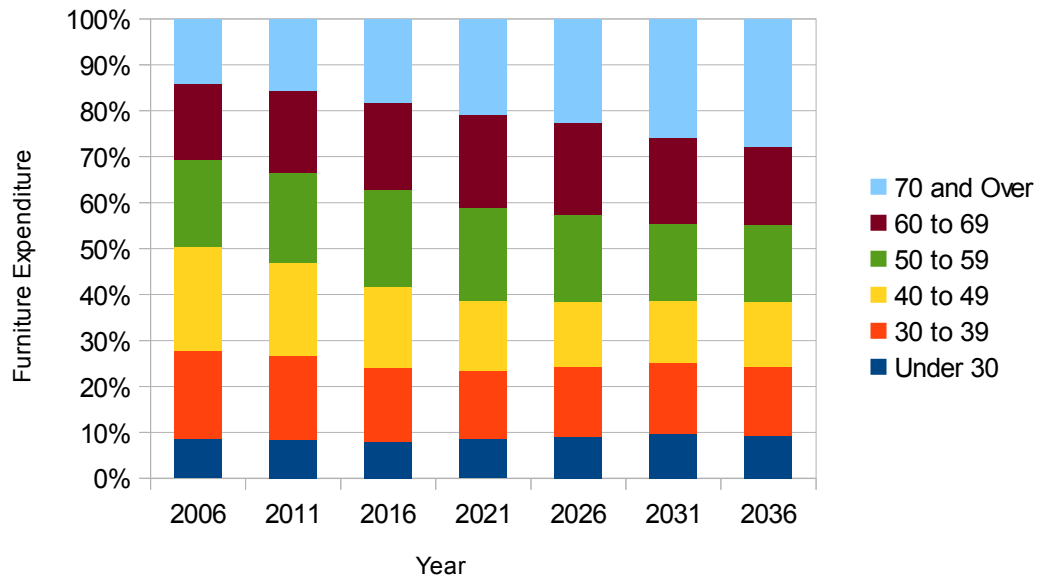


Figure 55: Share of Furniture Spending by Age Band

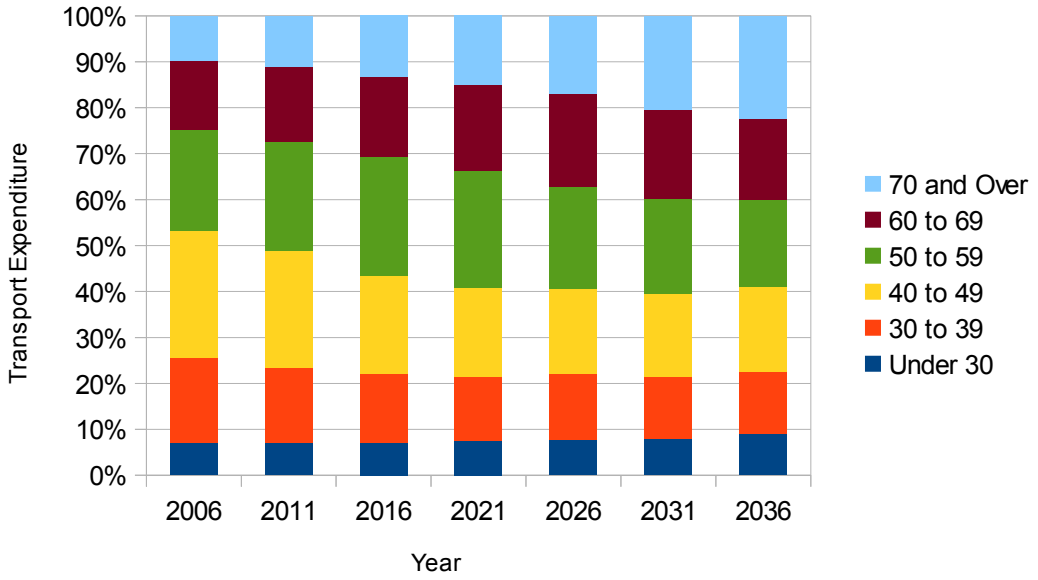


Figure 56: Share of Transport Spending by Age Band

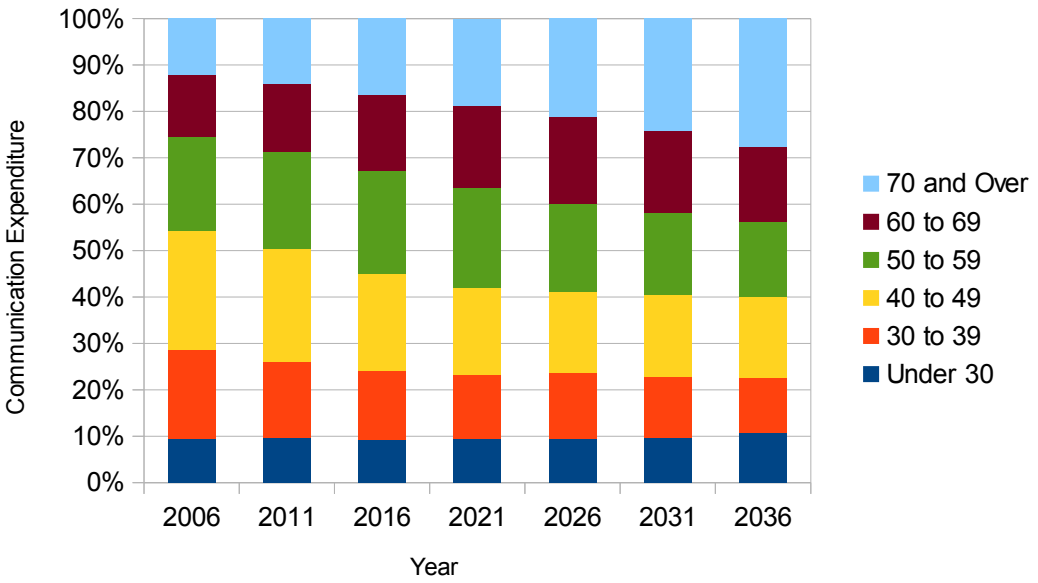


Figure 57: Share of Communication Spending by Age Band

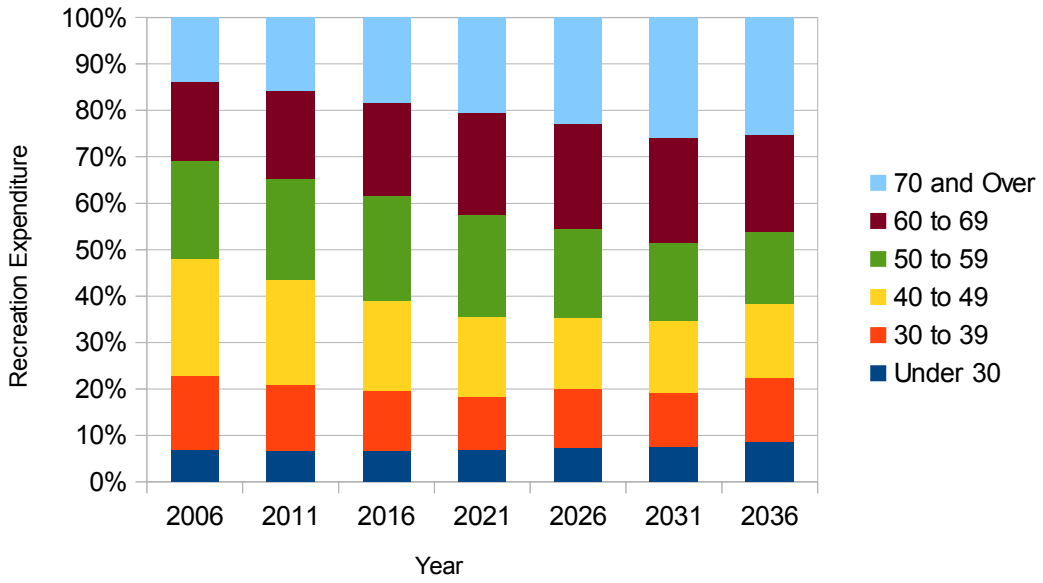


Figure 58: Share of Recreation Spending by Age Band

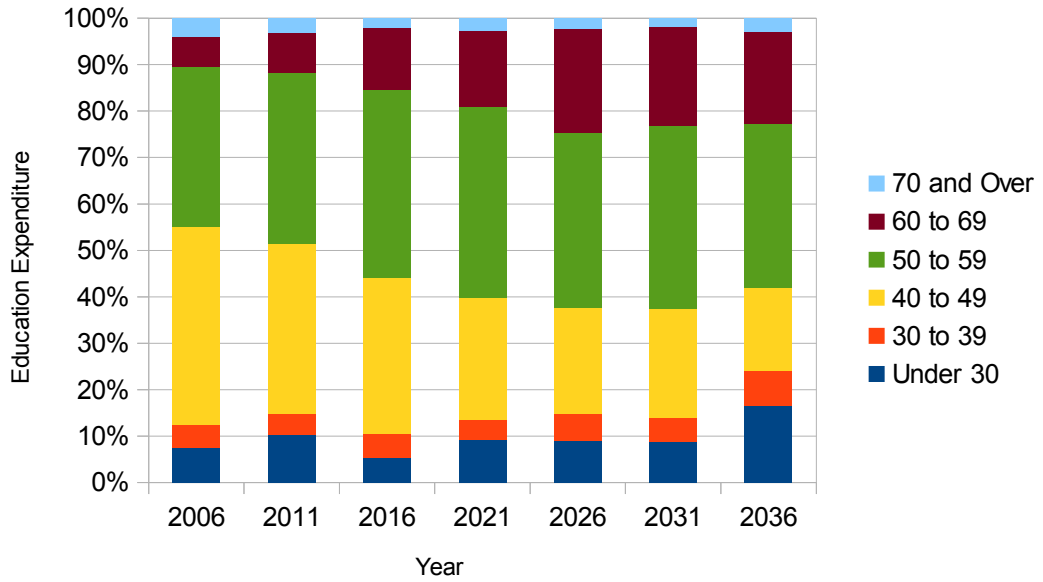


Figure 59: Share of Education Spending by Age Band

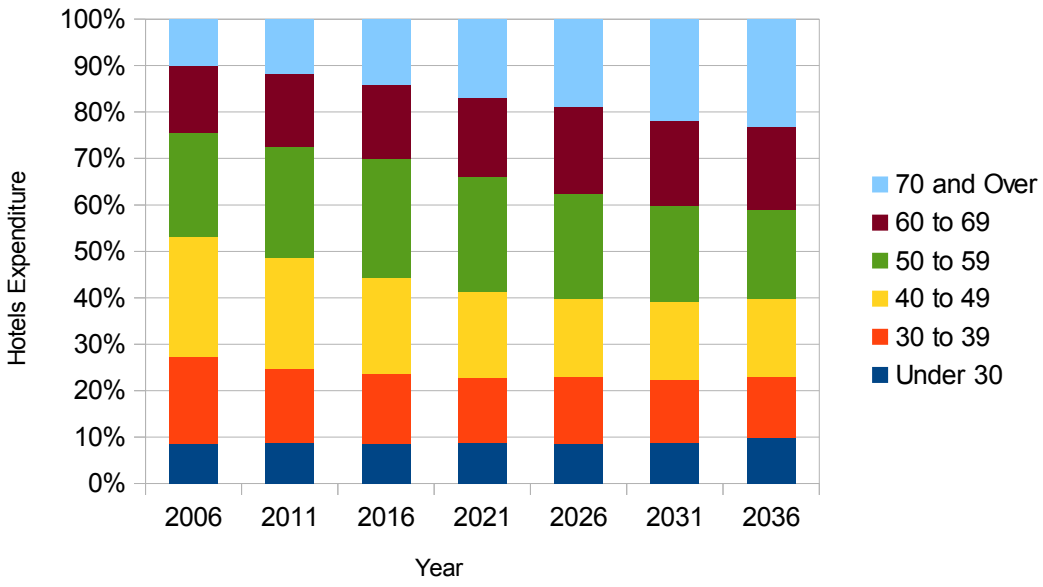


Figure 60: Share of Hotels Spending by Age Band

6.6.2 Spending by Household Type

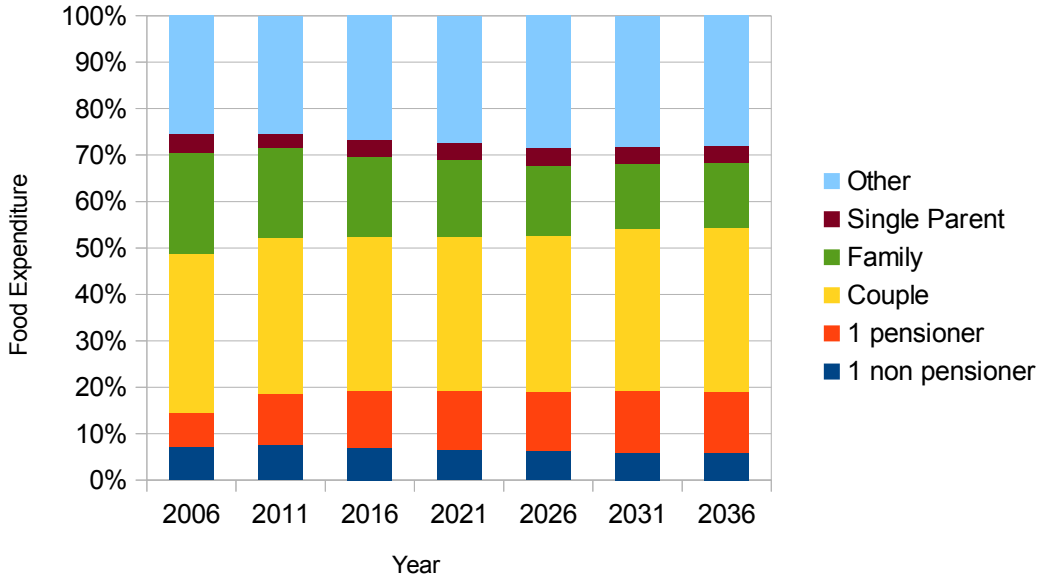


Figure 61: Share of Food Spending by Household Type

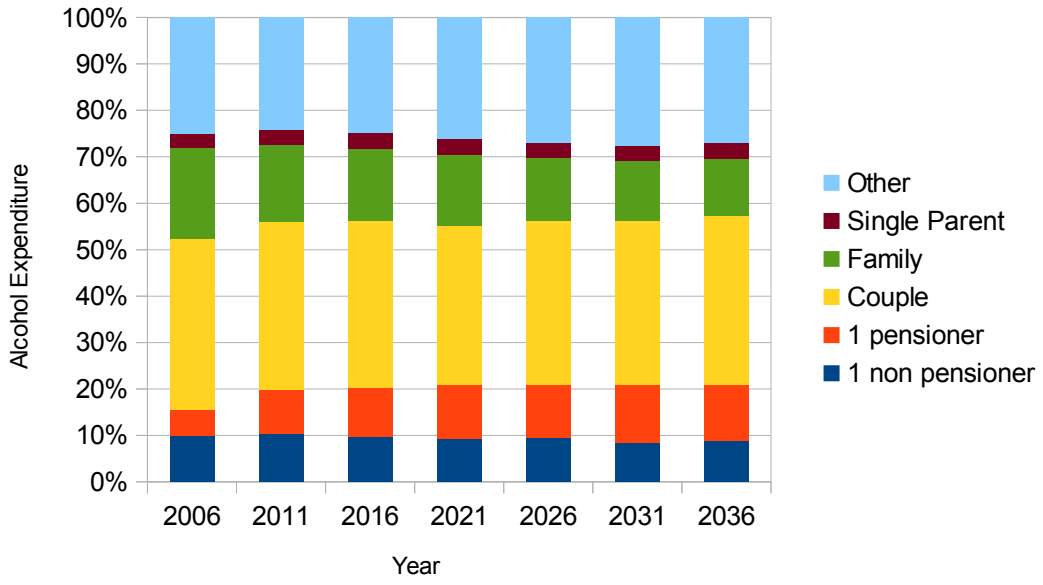


Figure 62: Share of Alcohol Spending by Household Type

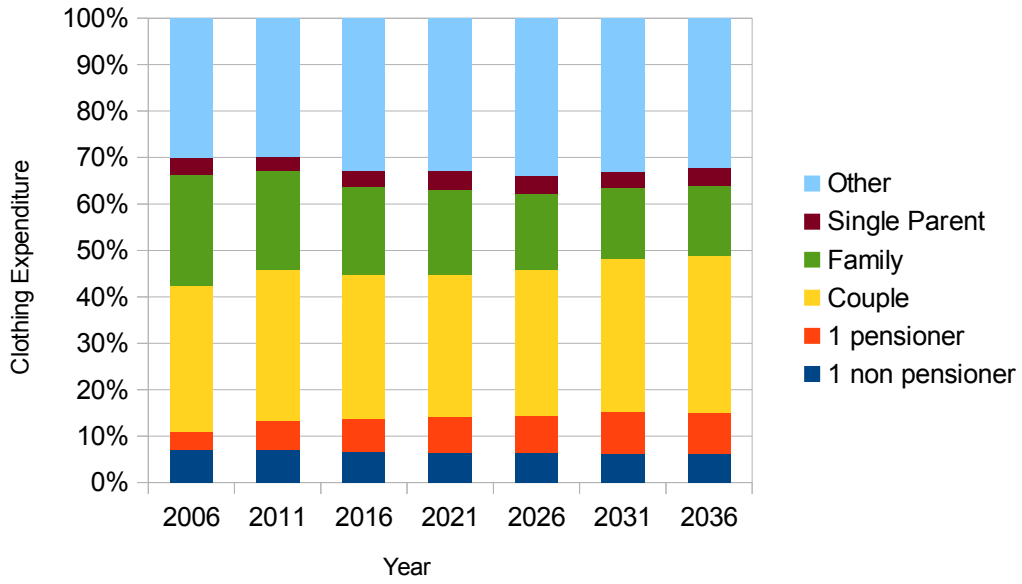


Figure 63: Share of Clothing Spending by Household Type

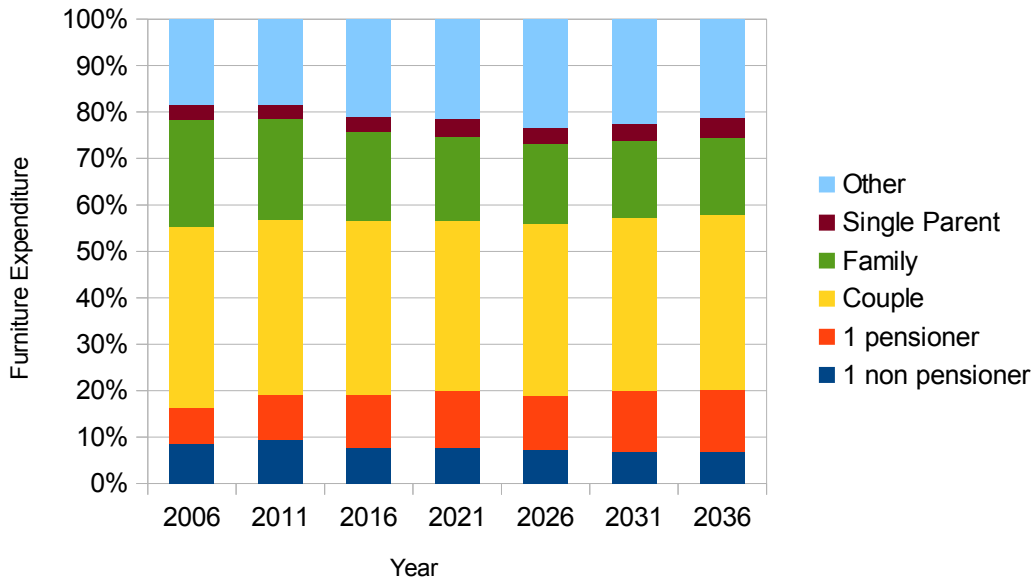


Figure 64: Share of Furniture Spending by Household Type

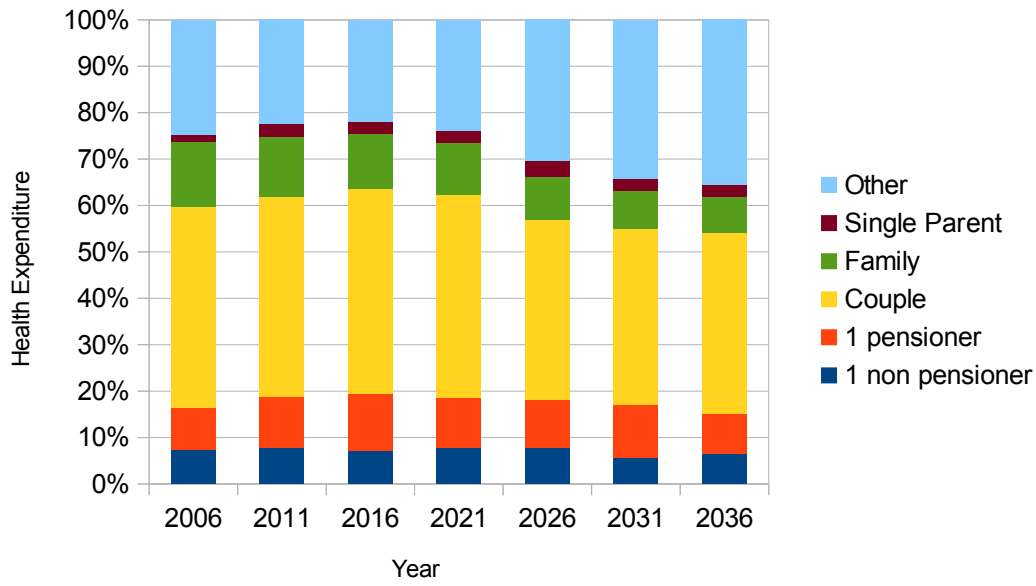


Figure 65: Share of Health Spending by Household Type

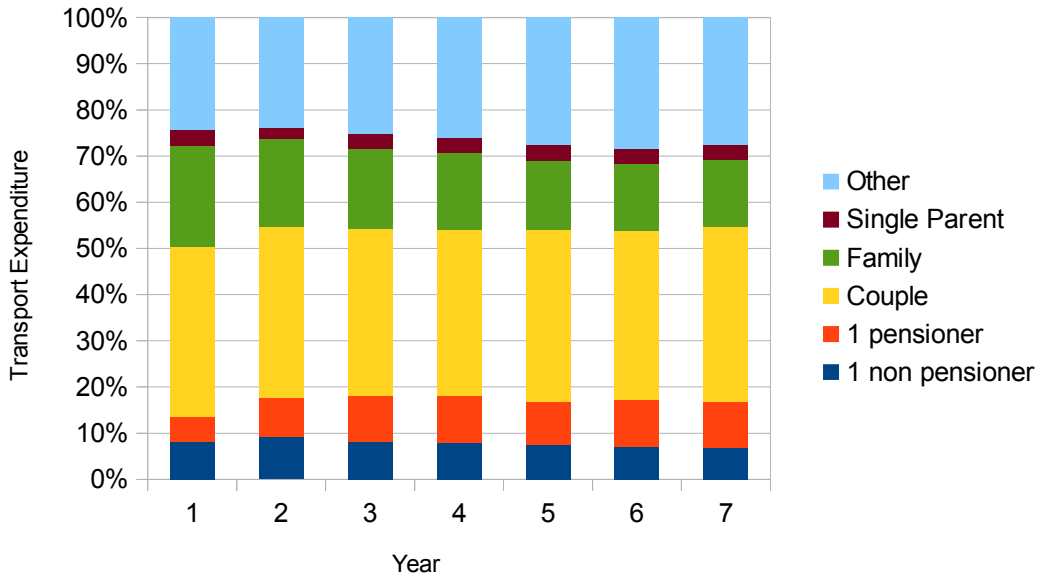


Figure 66: Share of Transport Spending by Household Type

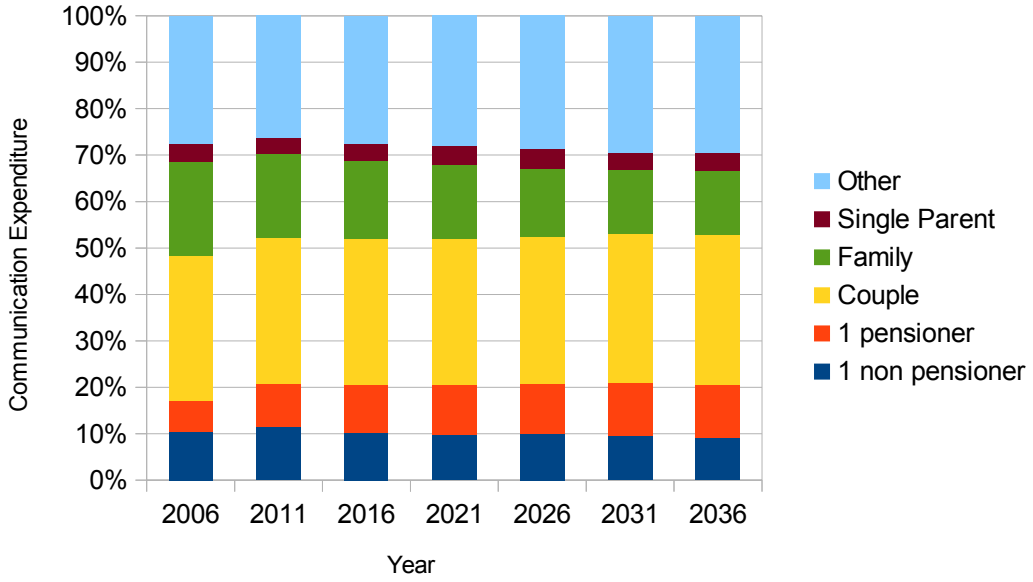


Figure 67: Share of Communication Spending by Household Type

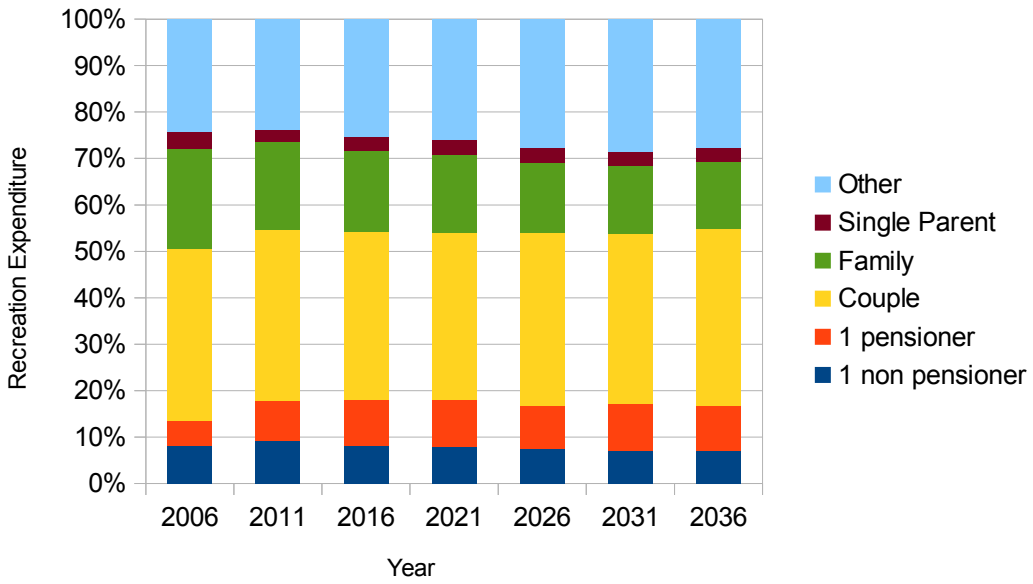


Figure 68: Share of Recreation Spending by Household Type

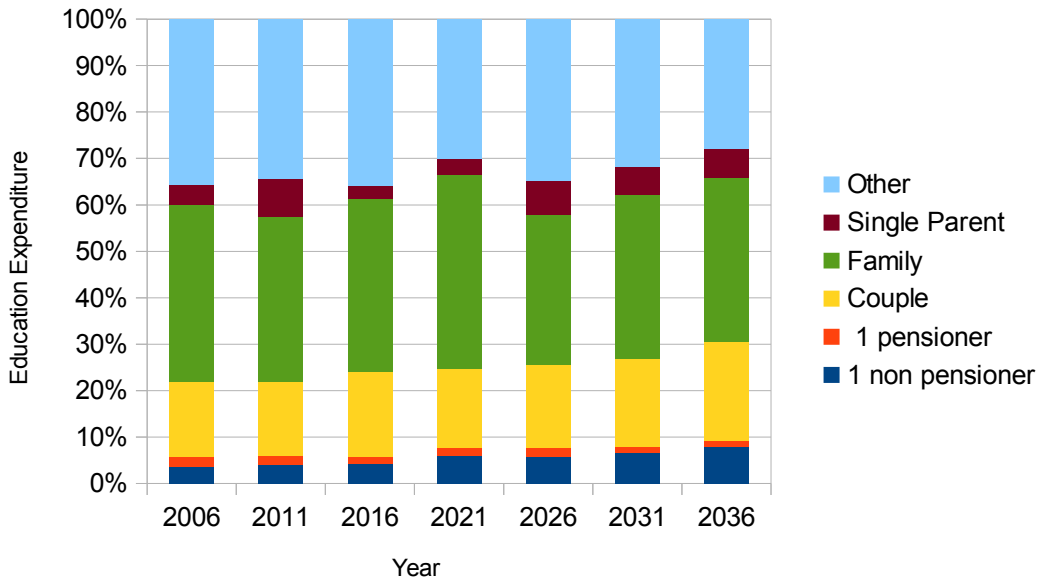


Figure 69: Share of Education Spending by Household Type

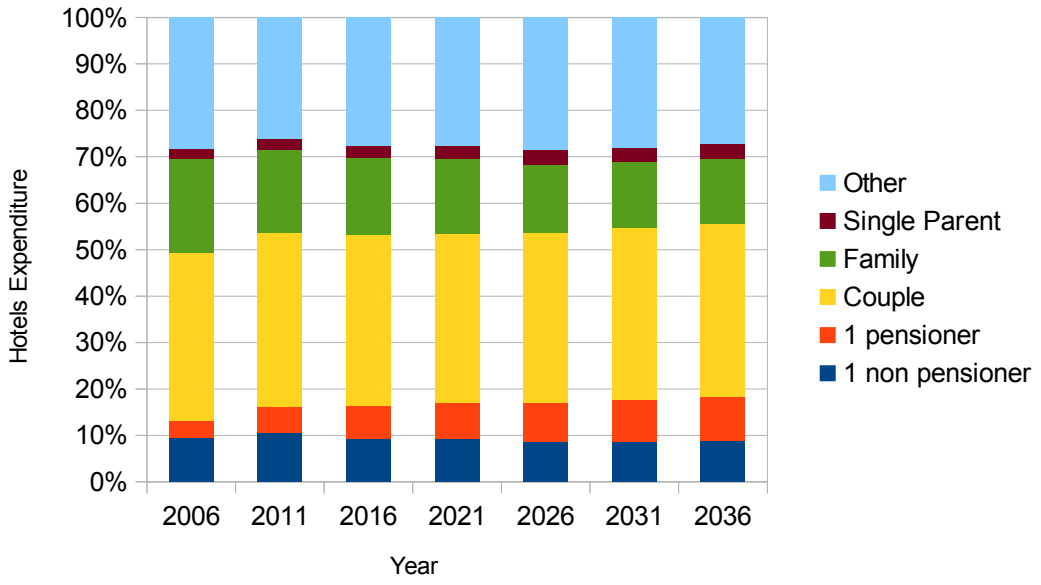


Figure 70: Share of Hotels Spending by Household Type

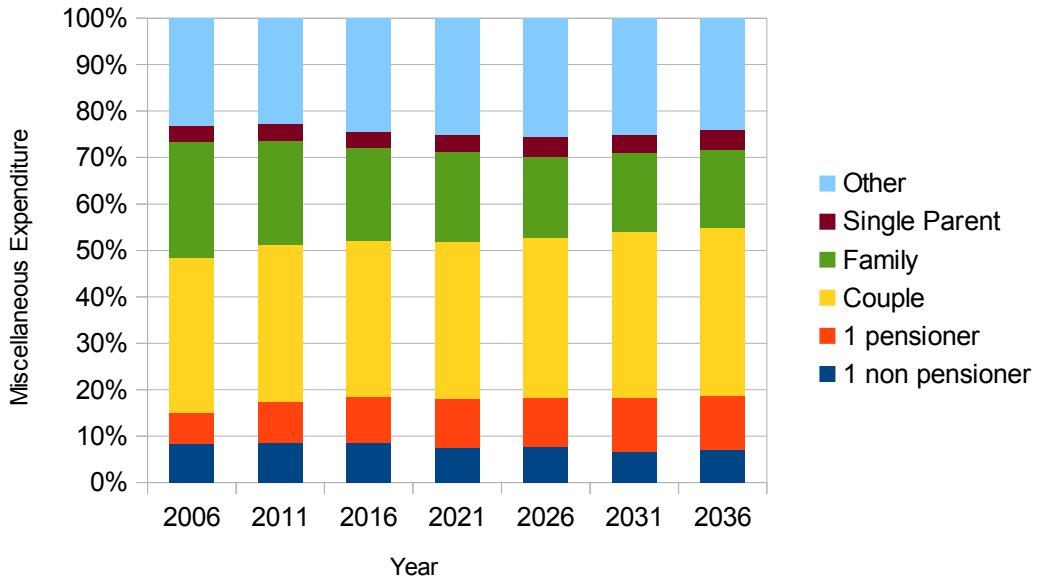


Figure 71: Share of Miscellaneous Spending by Household Type

6.7 Alternative Economic Scenarios

The previous results were based on the scenario of natural change in the base population. This was done to isolate demographic change from other factors in order to investigate the effect of population ageing on expenditure patterns. It is also possible to add further rounds of random assignment to superimpose the effects of other factors. In this section, an income change component, derived from the model in Chapter 5 is added to the combined demographic and random assignment model to simulate the effect of various combinations of demographic and economic change.

The pseudo-code for the complete model is given below.

```
set model parameters
load cross sectional household data file
load individual level data file
assign individuals to households using the household identification number (HID)
for each year
   for each household
      update demographic characteristics for 1 year
   for each household
      update expenditure pattern
   for each household
         while mean population income <= target population income
                  select uniformly distributed random number from 0 to 1
                  if number <= 0.5 (copy from next higher income household)
                           if any other households with (income >= target)
                                    and same type and number in household
                                             store donor expenditure pattern
                           else
```

```
                    change all spending categories in proportion to income
                                                                    change
        if number > 0.5 (copy from next lower income household)
            if any other households with (income <= target)
                    and same type and number in household
                        store donor expenditure pattern
            else
                    change all spending in proportion to income change
  for each household
      update expenditure patterns from stored donor
calculate new aggregate expenditures for goods of interest
```

Four different scenarios of population and income changed were investigated.

Scenario 1 – Population Change Only

Only natural change in the base population is considered. This is the same as the model described above.

Scenario 2 – Income Change Only

Household composition is kept constant while all households experience a 2% rise in real income every year.

Scenario 3 – Population Change and Rising Income

Scenarios 1 and 2 are combined to simulate the effect of rising incomes and population change. This is meant to represent the effect of a sustained period of economic prosperity.

Scenario 4 - Population Change and Falling Income

Population change is modelled and real incomes are reduced by 2% per year to represent a long period of economic austerity.

In all cases, the base population was obtained from the EFS as before and the simulation ran from 2006 to 2036. The results are summarised in the tables below and represent the average of 10 simulations. In Table 32, aggregate expenditure in 2006 is set to 100. The other columns (Scen. 1 to Scen. 4) show the percentage change from the base year in scenario 1 to 4 respectively. Discussion of the results is left until Section 6.8 below.

Category	2006 (%)	Scen. 1 Pop. Change Only	Scen. 1 95% Conf. (N = 10)	Scen. 2 Income Change (+2%)	Scen. 2 95% Conf. (N = 10)	Scen. 3 Pop. and Income Change (+2%)	Scen. 3 95% Conf. (N = 10)	Scen. 4 Pop. and Income Change (-2%)	Scen. 4 95% Conf. (N = 10)
Total Expenditure	**100**	**4.06**	6.07	**34.96**	36.29	**40.15**	43.74	**-24.7**	-23.43
			2.05		33.63		36.56		-25.97
Food	**100**	**10.26**	11.71	**14.01**	14.94	**26.44**	28.42	**-12.16**	-11.27
			8.81		13.01		24.46		-13.05
Alcohol	**100**	**8.25**	12.04	**7.69**	9.13	**17.25**	19.77	**-17.05**	-12.95
			4.46		6.25		14.73		-21.15
Clothing	**100**	**-1.98**	2.04	**28.92**	30.79	**30.03**	36.7	**-32.54**	-29.86
			-6		27.05		23.46		-35.22
Housing	**100**	**6.55**	10.63	**20.62**	21.96	**19.13**	22.93	**-2.53**	-0.61
			2.47		19.28		15.33		-4.45
Furniture	**100**	**4.24**	11.05	**30.62**	32.32	**39.18**	45.86	**-16.48**	-12.58
			-2.57		28.92		32.5		-20.38
Health	**100**	**16.45**	27.98	**14.97**	17.52	**46.34**	66.44	**-11.13**	-3.43
			4.92		12.42		26.24		-18.83
Transport	**100**	**5.2**	7.77	**46.86**	49.86	**53.36**	56.79	**-32.25**	-31.12
			2.63		43.86		49.93		-33.38
Communication	**100**	**5.53**	8.01	**14.15**	14.58	**21.3**	24.01	**-16.27**	-13.65
			3.05		13.72		18.59		-18.89
Recreation	**100**	**7.28**	9.12	**28.24**	29.44	**35.56**	38.29	**-23.02**	-21.3
			5.44		27.04		32.83		-24.74
Education	**100**	**-19.2**	-3.75	**274.66**	292.13	**249.8**	309.6	**-70.92**	-65.23
			-34.56		257.19		190		-76.61
Hotels	**100**	**4.74**	8	**37.89**	39.59	**43.53**	49.84	**-24.83**	-23.43
			1.48		36.19		37.22		-26.23
Miscellaneous	**100**	**4.27**	6.71	**30.23**	31.64	**35.45**	39.98	**-27.2**	-25.06
			1.83		28.82		30.92		-29.34
Other	**100**	**-1.36**	1.88	**41.24**	42.93	**43.54**	54.19	**-36.43**	-33.36
			-4.6		39.55		32.89		-39.5

Table 32: Aggregate Expenditure Change 2006 - 2036 in Four Scenarios

Table 33 below, shows the change in share of household expenditure under the four scenarios.

Category	2006 (%)	Scen. 1 Pop. Change Only	Scen. 1 95% Conf. (N = 10)	Scen. 2 Income Change (+2%)	Scen. 2 95% Conf. (N = 10)	Scen. 3 Pop. and Income Change (+2%)	Scen. 3 95% Conf. (N = 10)	Scen. 4 Pop. and Income Change (-2%)	Scen. 4 95% Conf. (N = 10)
Food	**9.86**	**6**	7.75	**-15.52**	-15.22	**-9.71**	-7.45	**16.71**	18.8
			4.24		-15.83		-11.97		14.61
Alcohol	**2.33**	**4.12**	7.39	**-20.45**	-19.51	**-16.21**	-14.73	**10.24**	16.03
			0.85		-21.4		-17.7		4.44
Clothing	**4.83**	**-5.88**	-3.77	**-4.78**	-3.64	**-7.08**	-2.3	**-10.43**	-7.2
			-8		-5.92		-11.86		-13.66
Housing	**9.83**	**2.4**	5.94	**-10.54**	-9.75	**-14.95**	-12.66	**29.49**	31.48
			-1.14		-11.32		-17.23		27.5
Furniture	**6.4**	**0.05**	5.26	**-3.41**	-2.35	**-0.69**	3.44	**10.81**	14.74
			-5.15		-4.47		-4.82		6.89
Health	**1.24**	**11.65**	22.37	**-14.18**	-11.96	**4.44**	18.4	**17.82**	27.2
			0.94		-16.4		-9.53		8.43
Transport	**12.97**	**1.14**	3.63	**9.05**	10.53	**9.54**	11.93	**-9.97**	-8.24
			-1.35		7.56		7.15		-11.71
Communication	**2.45**	**1.35**	3.84	**-15.34**	-14.6	**-13.26**	-10.3	**11.26**	15.16
			-1.14		-16.09		-16.23		7.37
Recreation	**12.33**	**3.13**	4.38	**-4.77**	-4.15	**-3.17**	-1.36	**2.21**	3.49
			1.89		-5.39		-4.99		0.93
Education	**1.48**	**-22.37**	-7.98	**179.84**	190.84	**147.88**	186.09	**-61.35**	-53.51
			-36.76		168.84		109.67		-69.2
Hotels	**7.95**	**0.64**	2.65	**2.1**	3.88	**2.67**	8.66	**-0.14**	1.52
			-1.37		0.32		-3.31		-1.8
Miscellaneous	**7.57**	**0.25**	2.58	**-3.6**	-3.03	**-3.26**	0.28	**-3.33**	-0.99
			-2.09		-4.18		-6.79		-5.66
Other	**20.88**	**-5.24**	-3.11	**4.37**	4.98	**2.21**	7.44	**-15.67**	-12.6
			-7.37		3.77		-3.02		-18.73

Table 33: Change in Budget Share 2006 to 2036 in Four Scenarios

6.7.1 Detailed Analysis of Scenario 2

The results for the scenario of population change and income increase of 2% per year are analysed in more detail below. This was done by reading the micro-level output from one simulation run into SPSS and executing case summaries for 2006 and 2036.

As discussed earlier, there is stochastic variation between simulation runs, which can be estimated from the distribution of values obtained from running the model several times. In consequence, these results in this section, derived from a single run, can only be indicative of the kind of disaggregated output that can be generated using random assignment. If the micro-level data file was obtained using multiple imputation (Rubin, 1987) then the level of uncertainty in the results could be allowed for. However, the size of the population is varying from year to year and there does not seem to be a way to utilise multiple imputation on varying file sizes. This problem could form a subject of further research.

Table 34 below, compares the distribution of food spending in 2006 and 2036. Total spending on food is broken down, first into age bands of the 'oldest person in the household', for the 'under 30' to 'over 70' groups, in ten year intervals as before. Within these categories, spending is distributed between the five income quintiles with quintile 1 being the lowest and quintile 5 being the highest. The number of cases is given, then mean household spending for each group is shown, followed by the share of total spending on food.

Food Spending in 2006

Case Summaries

food

Age of Oldest Occupant (Binned)	incanon (Binned)	N	Mean	% of Total Sum
Under 30	<= 243.4	445	24.3657	0.9%
	243.5 - 396.9	411	26.6563	0.9%
	397.0 - 627.6	506	31.5827	1.4%
	627.7 - 940.0	468	40.9318	1.6%
	940.1+	257	44.9981	1.0%
	Total	2087	32.8222	5.9%
30 - 39	<= 243.4	344	29.9608	0.9%
	243.5 - 396.9	752	35.0275	2.3%
	397.0 - 627.6	950	45.0937	3.7%
	627.7 - 940.0	1067	48.6995	4.5%
	940.1+	1120	54.1916	5.2%
	Total	4233	45.3917	16.5%
40 - 49	<= 243.4	471	27.6716	1.1%
	243.5 - 396.9	637	41.0526	2.2%
	397.0 - 627.6	1120	50.0438	4.8%
	627.7 - 940.0	1334	60.0483	6.9%
	940.1+	1510	71.1419	9.2%
	Total	5072	55.7495	24.3%
50 - 59	<= 243.4	703	30.0384	1.8%
	243.5 - 396.9	587	34.8288	1.8%
	397.0 - 627.6	835	51.8867	3.7%
	627.7 - 940.0	1123	57.2266	5.5%
	940.1+	1336	76.0759	8.7%
	Total	4584	54.7098	21.5%
60 - 69	<= 243.4	871	29.8815	2.2%
	243.5 - 396.9	931	40.5974	3.2%
	397.0 - 627.6	774	51.4494	3.4%
	627.7 - 940.0	617	57.7510	3.1%
	940.1+	545	71.2699	3.3%
	Total	3738	47.6510	15.3%
70+	<= 243.4	2137	26.8810	4.9%
	243.5 - 396.9	1653	38.2401	5.4%
	397.0 - 627.6	781	47.4002	3.2%
	627.7 - 940.0	360	59.0074	1.8%
	940.1+	198	74.0363	1.3%
	Total	5129	37.7417	16.6%
Total	<= 243.4	4971	27.9162	11.9%
	243.5 - 396.9	4971	37.1954	15.9%
	397.0 - 627.6	4966	47.3290	20.2%
	627.7 - 940.0	4969	54.8125	23.4%
	940.1+	4966	67.4229	28.7%
	Total	24843	46.9304	100.0%

Food Spending in 2036

Case Summaries

food

Age of Oldest Occupant (Binned)	incanon (Binned)	N	Mean	% of Total Sum
Under 30	<= 312.30	533	31.4830	1.2%
	312.31 - 455.02	436	39.5365	1.2%
	455.03 - 740.55	605	46.5425	1.9%
	740.56 - 1168.52	543	49.7081	1.9%
	1168.53+	462	66.7652	2.1%
	Total	2579	46.5349	8.3%
30 - 39	<= 312.30	382	29.8207	0.8%
	312.31 - 455.02	372	37.4226	1.0%
	455.03 - 740.55	604	45.9750	1.9%
	740.56 - 1168.52	979	53.3597	3.6%
	1168.53+	631	74.9602	3.3%
	Total	2968	51.4220	10.5%
40 - 49	<= 312.30	298	28.1130	0.6%
	312.31 - 455.02	335	35.5748	0.8%
	455.03 - 740.55	636	46.8893	2.1%
	740.56 - 1168.52	899	51.6760	3.2%
	1168.53+	1363	78.2901	7.4%
	Total	3531	57.5709	14.0%
50 - 59	<= 312.30	345	28.2981	0.7%
	312.31 - 455.02	377	38.9401	1.0%
	455.03 - 740.55	615	42.4101	1.8%
	740.56 - 1168.52	995	50.6584	3.5%
	1168.53+	1544	74.7813	8.0%
	Total	3876	55.8289	14.9%
60 - 69	<= 312.30	721	35.5526	1.8%
	312.31 - 455.02	1030	48.0206	3.4%
	455.03 - 740.55	1035	48.0921	3.4%
	740.56 - 1168.52	792	51.8189	2.8%
	1168.53+	919	82.3100	5.2%
	Total	4497	53.7144	16.7%
70+	<= 312.30	3314	38.4404	8.8%
	312.31 - 455.02	2989	47.7984	9.8%
	455.03 - 740.55	2116	48.0369	7.0%
	740.56 - 1168.52	1315	51.5160	4.7%
	1168.53+	639	120.7253	5.3%
	Total	10373	49.8211	35.6%
Total	<= 312.30	5593	35.6405	13.7%
	312.31 - 455.02	5539	45.1503	17.2%
	455.03 - 740.55	5611	46.9172	18.1%
	740.56 - 1168.52	5523	51.5800	19.6%
	1168.53+	5558	81.5228	31.2%
	Total	27824	52.1369	100.0%

Table 34: Changes in Spending for Food between 2006 and 2036

Table 35 below, focuses on food expenditure for the '70 and over' age group further disaggregated by household size as it was in 2006. Then, Table 36 shows how it is projected to be in 2036 for comparison.

Age Band (2006)	Income Quintile	Household Size	N	Mean	Percent of Total
70+	<= 243.4	1.00	1685	22.9855	3.3%
		2.00	428	40.6274	1.5%
		3.00	20	52.7240	0.1%
		4.00	4	67.8300	0.0%
		Total	2137	26.8810	4.9%
	243.5 - 396.9	1.00	669	24.7417	1.4%
		2.00	927	47.1515	3.7%
		3.00	49	56.7724	0.2%
		4.00	8	20.9150	0.0%
		Total	1653	38.2401	5.4%
	397.0 - 627.6	1.00	146	30.9762	0.4%
		2.00	513	49.8760	2.2%
		3.00	83	50.5504	0.4%
		4.00	39	69.6151	0.2%
		Total	781	47.4002	3.2%
	627.7 - 940.0	1.00	45	28.8756	0.1%
		2.00	204	53.8661	0.9%
		3.00	64	70.9358	0.4%
		4.00	47	93.9298	0.4%
		Total	360	59.0074	1.8%
	940.1+	1.00	7	17.0214	0.0%
		2.00	109	69.8745	0.7%
		3.00	36	78.7056	0.2%
		4.00	46	88.9198	0.4%
		Total	198	74.0363	1.3%
	Total	1.00	2552	23.9905	5.3%
		2.00	2181	48.2757	9.0%
		3.00	252	61.1321	1.3%
		4.00	144	80.9628	1.0%
		Total	5129	37.7417	16.6%

Table 35: Food Expenditure 2006

Age Band (2036)	Income Quintile	Household Size	N	Mean	Percent of Total
70+	<= 312.30	1.00	2003	36.8105	5.1%
		2.00	902	37.8383	2.4%
		3.00	236	52.4417	0.9%
		4.00	173	41.3500	0.5%
		Total	3314	38.4404	8.8%
	312.31 - 455.02	1.00	1317	48.3412	4.4%
		2.00	1125	48.8941	3.8%
		3.00	369	41.5572	1.1%
		4.00	178	49.7956	0.6%
		Total	2989	47.7984	9.8%
	455.03 - 740.55	1.00	798	46.2824	2.5%
		2.00	1063	47.2553	3.5%
		3.00	210	58.7122	0.8%
		4.00	45	47.7940	0.1%
		Total	2116	48.0369	7.0%
	740.56 - 1168.52	1.00	301	52.2317	1.1%
		2.00	610	48.5717	2.0%
		3.00	353	54.9991	1.3%
		4.00	51	58.3992	0.2%
		Total	1315	51.5160	4.7%
	1168.53+	1.00	126	71.0337	0.6%
		2.00	349	118.6168	2.9%
		3.00	53	313.2179	1.1%
		4.00	111	91.8505	0.7%
		Total	639	120.7253	5.3%
	Total	1.00	4545	43.7849	13.7%
		2.00	4049	51.9621	14.5%
		3.00	1221	62.2896	5.2%
		4.00	558	56.1679	2.2%
		Total	10373	49.8211	35.6%

Table 36: Food Expenditure 2036

6.8 Discussion

The modelling work described in this chapter has two aims. The first is the substantive task of obtaining and interpreting projections of the effect of population ageing on household expenditure patterns. The second is to evaluate random

assignment and NetLogo as applied in a 'real world' application. The following discussion reflects this and is divided into two parts. Section 6.8.1 considers the results of the projections and uses them to make inferences to spending behaviour at the macroeconomic level. Following that, Section 6.8.2 discusses what new information this reveals about using a random assignment scheme to model household expenditure.

6.8.1 Comments on the results

These results provide some examples of the way the micro-level data that is generated by the microsimulation can be aggregated and presented. The projections for the first sections are intended to show the effect of natural demographic change in isolation from other factors so that the component of spending patterns, which is attributable to the ageing population, can be identified and quantified.

The results began with Figure 29 which showed that average household income decreases between 2006 and 2036. As the population ages, the proportion of pensioners increases and since this group tend to have a lower income than they did when they were of working age, this has a moderating effect on incomes overall. The next graph shows that total aggregate income for the whole population rises between 2006 and 2036. This is due to increasing life expectancy leading to pensioners drawing their pension for a longer period than they would otherwise have done so. Several microsimulation models have calculated the additional cost of pensions in an ageing population. This model shows how the costs manifest in greater aggregate

income and then to changes in spending patterns. Table 30 indicated that 'total expenditure' rises 5.5% by 2021 and then reverts to a 2.1% increase by 2036, compared to the 2006 baseline. According to the ONS (2007b), total UK domestic expenditure in 2006 was £782.809 billion. As a result, it is possible to infer that population ageing adds *ceteris paribus* £43.05 billion (5.5% of 782.809 billion) to UK total consumption expenditure in the year 2021. Over the period 2006 to 2031, the average increase in total expenditure is 3.5% or 27.4 billion per year which entails a total extra spend of 685 billion over 25 years. In 2006, UK GDP was 1,328.363 billion (Hills, Thomas and Dimsdale, 2010) so population ageing adds an amount which is numerically equivalent to ((27.4 / 1,328.363) * 100) = 1.98% of GDP over the 30 year period; all in 2006 prices. The source of this extra spending is the additional pension income of the larger group of older people who, due to increasing life expectancy survive for longer.

It seems clear that this amount of additional expenditure, which is attributable to population ageing, would have a significant effect on the wider UK economy. One factor that is apparent, is that the indirect taxes paid on the extra consumption would go some way to offset the increases in healthcare and pension costs already projected in microsimulation studies. It is also possible that the increased demand may provide a boost to the economy by facilitating increased output and employment. However, if domestic output does not satisfy the extra demand, there is a risk of a rise in inflation or imports. A precise determination of the economic effect would require further research but the results produced here show something that has rarely been mentioned

in previous research which is that there is at least the possibility that the ageing population might bring some economic benefit.

Several of the other COICOP categories show a similar pattern to what is projected for 'total expenditure', with modest rises to a peak: 'hotels' (4.3%), 'housing' (2.2%), 'furniture' (7.4%), and 'miscellaneous' (6.4%), while 'food' (10%) and 'transport' (8.5%) have slightly larger increases. The largest rise is for 'health' which peaks at 31.7% above its 2006 baseline before reducing to 17.4% above 2006 values by 2036. 'Education' rises to 11% above its 2006 baseline in 2016 before declining to 85.6% in 2036. 'Clothing' at 98.4% in 2036 is the only other expenditure category to fall below its starting value.

After the graphs showing aggregate expenditure for the COIOCOP categories, results were presented for 'housing', 'health', 'transport' and 'food', disaggregated into spending by age bands. These indicated that the additional expenditure, induced by population ageing, is not spread evenly by age groups. The majority comes from an increase in the number of households in which the oldest person is aged 70 and over. This more than compensates for a decline in spending attributable to households in the '30 to 39', '40 to 49' and '50 to 59' age bands. However, due to decreased household size and the reduced spending power of the mostly retired, over 70s, average spending per household in the UK overall falls, while aggregate expenditures are rising. This is manifest in the observation of a 2.1% rise in 'total expenditure' by 2036, by which time the number of households rises by 12.7%.

Detailed inspection of how spending patterns change according to age group can shed some light on the source of the changes in aggregate expenditure over time. Here, the model is used for explanation rather than prediction. In Figure 38, aggregate spending on 'health' appeared to decrease towards the end of the series. This might seem surprising when the population continued to age over the whole of the period considered. Figure 48 shows expenditure for 'health' by age band and indicates that spending in this category, while continuing to rise for the '70 and over' age group, started to decrease for the '30 to 39', '40 to 49' and '60 to 69' age bands. This is particularly evident for the '60 to 69' age group where spending, after initial rises, begins to decline after 2021.

In contrast to 'health', spending on 'housing' shows a slight rise towards the end of the series. Inspection of Figure 47 shows that the early declines in spending for this category level off for the '30 to 39', '40 to 49' and 50 to 59' age groups towards the end of the series, while spending by the '70 and older' group keeps increasing. Overall, this is sufficient to maintain the increased level of spending on 'housing'.

While the EFS only provides data at the individual and household level, aggregating across the COICOP categories provides an indication of the level of demand within the corresponding industrial sector. This can provide some insight into how the characteristics of customers for particular types of commodity vary over time. Analysis of spending by industrial sector, shown in figures 51 to 73, indicates that the marked changes in expenditure by age band, translate into moderate shifts in demand

within sectors. This is partly because the spending power of the numerically larger group of people aged 70 and over, is attenuated by their decreased income compared to the middle-aged section of the population. The graphs show a similar general pattern where expenditure by the '70 and over' age group displaces spending by others as a proportion of total spending. This is most pronounced for the '40 to 49' and '50 to 59' age bands while the 'under 30' households are less affected. This indicates that for many of the sectors, some of the current emphasis on middle-aged consumers will shift towards older people. This is most apparent in the sectors of 'food' and 'health' (Figures 51 and 56) where by 2036, more than half of spending will be attributable to households in which the oldest person is aged 60 and over – all other things being equal.

The model described here is based on a number of assumptions. One is for an increasing life expectancy while all other transition probabilities remain as they were in the period 1991 to 2006 (except for birth which rises slightly then reduces again). There is also a limitation in that the copying process implicitly assumes that the behaviour of households, given their demographic characteristics, remains constant over time. Also, there has been no attempt to include the effect of migration which according to ONS projections of population size (ONS, 2011b), is likely to have as much effect as the natural change modelled here. Nor is there any representation of the effect of economic factors such as changes in prices, income, interest rates, technology or changing tastes. The result of this is to isolate the effect of population ageing in the absence of other confounding factors. While not providing a forecast of

what is likely to happen in the future, these results contribute to understanding the economic effects of population ageing and show that increasing life expectancy leads to an increase in the UK population and so to a rise in demand for most expenditure categories. This extra spending has the potential to provide a boost to the UK economy, provided that demand can be accommodated by increased domestic output. In many of the sectors considered, this would seem to be plausible, given that the relatively moderate changes, in percentage terms, take place over a period of several years. However, the increased demand for housing which is indicated by the projected increase of 12.7% in the number of households may be more difficult to satisfy due to factors such as the limited availability of land, planning restrictions and the length of time needed to build new houses. The number of households in 2006 was calculated above to be 25,610,710 so a 12.7% increase represents the creation of about 3¼ million additional households. This may present one of the most significant challenges associated with population ageing.

Section 6.7 relaxed one of the model assumptions in order to investigate the effect of a combination of population and income change. Scenario 1 represents population change only, Scenario 2 showed the effect of steadily rising incomes. Scenario 3 showed the result of combining population change and rising incomes, Scenario 4 showed the effect of population change combined with a long-term decline in real household incomes. Table 32 shows that population change alone adds slightly over 4% to aggregate expenditure while a modest rise in incomes of 2% per year adds nearly 35% to overall spending. A combination of population change and income rise

leads to a 40% increase in spending between 2006 and 2036.

This pattern is not uniform across all spending categories. Aggregate 'food' spending rises 10% due to population change and 15% due to rising incomes. 'Alcohol' spending rises 8.25% due to population ageing but only 7.69% with rising incomes, ignoring demographic change. This is due to the declining budget share of 'alcohol' with rising incomes (shown in Table 28 in Section 5.4.1). Another notable feature of Table 32 is that spending on 'housing' only declines by 2.53% when allowing for population change and falling incomes. This is due to the combination of an increase in the number of households and the essential nature of this category. 'Education' appears to rise by 275% however the confidence intervals are also very large, indicating that this result should be treated with caution and may be due to the effect of a small number of unusual cases. Further investigation would be required to explain this result fully.

Tables 35 and 36 provided an illustration of how more disaggregated results can be obtained using the random assignment method. The example was to show how the distribution of spending for food would change between the year 2006 and 2036. This allowed a more detailed inspection of how the nature of consumers of this good would change over time taking into account both population and rising income. The final row for Table 35 shows that average spending on food for all households increased from £46.93 to £52.13 or slightly over 11%. The 70 and over age group increased their share of total spending on food from 16.6% in 2006 to 36.6% in 2036.

From the mean average column, it appears that a higher proportion of households headed by someone who is over 70 are in the top income quintile. This has risen from 198 to 639 (223%). Meanwhile, the number in the lowest income quintile increases from 4971 to 5593 (13%). Hence, it appears that a combination of the increased numbers and higher spending of the older group that accounts for their sharp increase in the share to total food consumption.

It can be seen from the totals breakdown towards the bottom of the table, that the highest and lowest income quintiles increase their spending as would be expected in this scenario of increasing incomes. However, the middle quintile appear to fractionally decrease average expenditure for food. This result appears to be unlikely and may be due to artefacts of the copying mechanism at the upper and lower ends of the income distribution where there is more variation between cases. Further investigation would be required to determine the source of this anomaly and how the copying procedure could be modified to allow for this effect.

The final table demonstrates a further level of disaggregation. Spending for 'food' (one of the 12 COICOP categories) is broken down by age band (6 categories), by income quintile (5 categories) and by household size (4 categories). If all 12 spending categories were presented at this level of detail, there would be 1440 categories. These results can only be indicative of the kind of output that can be obtained using dynamic microsimulation and random assignment because they were obtained from an analysis of the micro-level datafile from one run. It would be possible to run the

simulation several times to determine confidence intervals as was done with the earlier models. However, each of the variables would have to be coded into the program, assigned values and have the results calculated. This process becomes unwieldy with such a large number of variables. One way to solve this problem would be to conduct a multiple imputation so that a few micro-level output files could be combined and correct estimators could be obtained. The problem here is that different numbers of households are created in different runs and it is not clear how to multiply impute using different sized files. Further research in this area may be useful.

6.8.2 Comments on the Method

One feature that is noticeable from these results is the relatively large variation between simulations, indicated by the width of the 95% confidence intervals in Figures 32 to 44. They are relatively small for 'food' and 'total expenditure' but for 'health' the range is larger than the projected increase in expenditure itself. The source of the variation can be traced to the stochastic nature of the random assignment algorithm. As the program steps through cases, each household copies the expenditure pattern from a similar donor. This consists of a set of values representing the expenditure for each good. If one particular good is considered, every time a household is updated, this may be higher or lower than the current value and so increases or decreases total spending for the population. This disturbance can be thought of as having two components, a deterministic element due to the systematic difference between the characteristics of donor and recipient households and a

random element due to idiosyncratic differences which are not due to the observed differences between households. As the random disturbances accumulate, the total expenditure for the population moves in a series of small changes in a random direction superimposed on a steady movement in one direction. This can be characterised as a type of Markov process known as a 'random walk with drift' (Harvey, 2012). As noted in Chapter 2, microsimulation is also a type of Markov process and the similarity is a consequence of using the methods and concepts of microsimulation throughout.

In a random walk, the distance travelled d, from the starting point is given by $l\sqrt{N}$ where l is the average step size and N is the number of steps taken. In the model, the number of steps after one simulated year is equal to the number of households (one step for each household) and the average step size is the standard deviation of expenditures (the root mean square of the differences in expenditure from the mean). If the standard deviation is divided by the mean, then multiplied by 100, the resulting 'coefficient of variation' gives a measure of the spread of the parameter as a percentage, regardless of its magnitude. For expenditure on 'health' this is 580% and for 'education' it is 772% while for 'total expenditure' it is 88% and for 'food' it is 130%. This shows that the intrinsic variation in spending for 'health' and 'education' are much larger than those for 'total expenditure' and 'food' and this explains the difference in the range of the confidence intervals.

Much of the variation is due to the highly skewed distribution of expenditure patterns

within the UK population. Inspection of the EFS for 2006 shows that, for most items, the majority of households report either zero or a low level of expenditure during the two week diary report. In the 'health' category for example, about half of households report no spending and only 10% of households spend more than £10 per week. Meanwhile, a small number of households report very high expenditures, with the top five households spending: £601, £603, 658, £700 and £1,975 per week on health. The random assignment scheme operates by copying the expenditures from a similar household and therefore, the number of times the high spending households are selected will make a significant difference to the progress of each simulation.

Luhrmann dealt with extreme cases by excluding 'about 1% of the households because of outliers in total expenditure or expenditure shares for non-food items above 80 percent' (Luhrmann, 2008: 5). Here, in contrast to this approach, all values have been used as reported in the EFS. This was done for several reasons. First, the EFS is a professionally run survey where according to the Quality and Methodology Information Paper (ONS, 2012) questions are piloted to ensure that respondents understand them correctly and a process of quality assurance is applied during and after the data collection process. Second, a similar pattern is seen for all the expenditure groups. If the extreme values were due to errors or mistakes, it is unlikely that the pattern would be so consistent. Third, the ONS treats suspicious values for *income* by investigating and correcting errors where they are found. However, the variable 'incanon' (total income plus allowances) shows a similar pattern with top five values of: £7,788, £10,848, £12,692, 16,437 and £47,117 per week. Fourth, in a

snapshot survey such as the EFS, it is possible to have one-off payments which exceed usual weekly income appearing in the spending diary. Fifth, 'fat-tails' of extreme values for income and wealth have been observed many times by such as Pareto (1897) and Brzezinski (2013) so it is not surprising that similar unusual values appear in expenditure patterns as well. Sixth and finally, since there is no more reliable source of information available that could be used to supplant what is reported in the EFS, applying a common sense correction to reported values risks introducing more errors than it would remove.

The large confidence intervals did not appear in the income model of Chapter 5. This is because the ratio of drift to random variation is larger for income than for demographic change. It can be seen from Table 30 that demographic change leads to a 5.5% rise in 'total expenditure' by 2021 while the income scenario depicted a situation where there was a 37% increase in 'total expenditure'. In both cases however, the step size is the standard deviation of 'total expenditure' which is the same in both models initially and only increases by a 1:1 ratio as 'total expenditure' rises in the income model.

6.9 Conclusion

This chapter has presented an example of the application of a random assignment model using NetLogo to investigate the effect of the ageing population on household expenditure patterns. Much of the research into population ageing has focused on determining the cost of this phenomenon and the debate has usually centred around

whether or not these costs are affordable and what policy changes may be needed to accommodate them. The research described in this chapter indicates that while the creation of sufficient housing is a concern, population ageing may increase aggregate UK expenditure and potentially have some beneficial effects for the economy as a whole. As the 'manageability' perspective suggests, the ageing population will have numerous effects, which will not be limited to the additional costs of healthcare and pensions. All aspects of this phenomenon are needed for a proper understanding of the effects of population ageing.

While these substantive results are significant in themselves, the methods developed in this thesis have the wider aim of creating a coherent micro-level framework that has general applicability to a wide range of problems. The results presented in this chapter represent a case study of the application of random assignment and NetLogo to study the effects of an ageing population on household expenditure patterns. It showed that the combination has the advantages of being able to project the most relevant variables for the expenditure system and then utilise the full heterogeneity of households, including unobserved heterogeneity, to model change over time. This exercise demonstrates that the approach has sufficient flexibility to tackle a complex problem and adds to what was learned about random assignment and NetLogo in previous chapters.

Table 2 in Chapter 3 indicated that random assignment is limited to cases that have already been observed while demand system modelling is highly flexible in that it can

represent any kind of rational optimising behaviour. The next chapter begins to address this limitation of random assignment in two models that were developed in collaboration with the co-sponsor of this project, BT.

Chapter 7: Further Applications

7.1 Introduction

At the start of this research, it was anticipated that the methods developed would be put to use by the co-sponsor to contribute to the process of long term infrastructure and technology planning. This is to be done by developing a range of scenarios for the model to run and assessing the effect of social, economic and technological change on expenditure patterns for a range of goods including telecommunications.

The scenarios were formulated during a series of meetings and workshops with members of the co-sponsor's modelling team. These took place at intervals throughout the early and mid-part of the research but were particularly concentrated towards the end of years 1 and 2. Some of the considerations that went into shaping the nature of the scenarios were the requirements of the sponsor and the compatibility of the scenarios with the wider research aims of the project, as described in Chapter 1.

The sponsors were particularly interested in modelling the decision points for consumers and how these would relate to the demand for telecommunications. These included the choice between spending on transport and broadband. Also, the question of how spending on entertainment, especially internet based activities, would rank in priority when compared with other expenditure items and which type of expenditure items would be cut back if household finances were limited or reduced? The research aims of the project were, as stated in Chapter 1, to model household expenditure at a

disaggregated level allowing for demographic change and the heterogeneity of households.

Discussion against the background of these diverse requirements eventually converged around the idea of what would happen if limited natural resources led to rising energy prices. The objective was to model and forecast how the demand for telecommunications would be affected by rising energy prices and whether such changes might represent an opportunity for broadband services to perform some of the functions remotely, that were previously done by travelling to a certain location.

These requirements led to the development of two models. The first one is a modified and extended version of the income change model described in Chapter 5. This is designed to shed some light on questions like: if energy prices rise, what sectors of household spending will be affected and what will be the effect on telecommunications spending? The second model arises from the proposition that rising energy prices might lead to more people working from home, supported by advances in telecommunications technology, rather than commuting daily to a fixed place of work. Although some households might make significant savings from not travelling to work, this could be offset to some extent by extra heating and utility costs. The model is designed to estimate the financial effects of a shift in the the location of the workplace to find out whether the anticipated savings in travel costs would be realised in practice.

These two models are described in this chapter. Each begins with a brief introduction to the background of the problem. This is followed by a description of how the model works. Some results are presented, followed by a discussion of what is uncovered about the random assignment scheme. The conclusion to the chapter reviews the models, assesses their limitations and proposes ways in which they could be developed further.

7.2 Scenario 1: The Effect of Energy Prices on Household Spending

The supply and pricing of energy are of interest to the co-sponsor because the share of household spending on telecommunications has to compete with spending on other items such as transport, food and utilities, which are themselves affected by the price of energy. It has also been posited, during discussions with the co-sponsor, that a decrease in travelling brought about by increasing fuel costs might present opportunities for the development and application of new forms of telecommunications technology.

Energy pricing and its effects are of particular relevance today because of concerns surrounding three global economic factors. The first of these is the increasing *demand* for energy. This is not only the case, as it has been historically, in the developed world but it is increasingly in developing nations that the rising demand for energy is most evident (Doggett & Ayesha, 2009). If significant new sources of energy are not found, this process can only result in higher energy prices. The second major influence is *climate change*. While this does not have a significant effect on energy prices directly,

the response to global warming from governments and individuals can. Proposed 'green taxes', designed to reduce the consumption of CO2 producing energy sources would, by design, tend to increase the cost of energy. Also, a shift to alternative energy generating technologies may well be more expensive than current methods. Conversely, it is possible that more efficient energy utilisation brought about by the higher energy costs might offset some of the rise in energy prices. The third global factor that could affect energy prices arises from the fact that the *supply* of fuels for energy generation is finite. This is most acute in the case of oil (Hubbert, 1956) but in the longer term, supplies of coal will dwindle as will the sources of the uranium that is used in nuclear power generation (Dittmar, 2009). This, coupled with rising demand for energy practically guarantees turbulent times for energy generation in the decades ahead.

The severity and imminence of all these factors is the subject of much debate and it is beyond the scope of this thesis to attempt to predict what energy prices will be in a number of years from now. Rather, the aim of the models described in this chapter is to gain some understanding of the way that households would alter their spending patterns in response to changes in the cost of energy. It is hoped that this will make a contribution towards the co-sponsor's planning of the development of infrastructure and technology whatever energy price changes actually occur in the future.

7.2.1 Energy Price Change Model

Each sector of the economy has its own characteristic demand for energy. Some, for

example, power generation, are very energy intensive. Others, such as financial services use less energy. The co-sponsor has conducted an extensive study of the energy requirements of all sectors of the economy as defined by the Standard Industrial Classification (SIC) (Potter, 2010). In the computer model, expenditures are represented by the high level COICOP (Classification of Individual Consumption by Purpose) categories as used in the EFS. To determine what effect energy price changes have on spending within these categories, it is necessary to form some kind of mapping between the SIC categories and the COICOP codes. This was obtained form a spreadsheet prepared by Morgan Potter. He began by collecting data on the cost of energy within each sub-category of the SIC. This creates a measure called the 'sensitivity' which is the percentage of the cost of producing goods in that sector that is attributable to energy. Next, each COICOP category is decomposed into the SIC sub-categories that contribute to the cost of the commodity. A weighted average of these forms the COICOP sensitivity which indicates the percentage of the price for the good that is attributable to energy.

EFS Household Expenditure Category	COICOP Sensitivity (%)
Food & non-alcoholic Drinks	5.98
Alcohol Tobacco & Narcotics	5.51
Clothing & Footwear	4.33
Housing Fuel & Power	33.27
Household Furnishings & Equipment	19.46
Health	4.03
Transport	15.08
Communications	3.68
Recreation & Culture	11.16
Education	1.17
Restaurants & Hotels	2.71
Miscellaneous	9.35

Table 37: Sensitivity of Categories of Household Good to the Price of Energy

Table 37 shows, for each of the high level EFS categories, what percentage of the price is due to energy. This allows the cost (or saving) of a change in the price of energy to be calculated for each household depending on the amount they currently spend on each commodity. This is implemented by the formula.

$$\Delta E = \text{sensitivity}/100 * \text{current expenditure} * \text{energy price change}/100$$

Where ΔE is the change in spending on the commodity in pounds per week.

Current expenditure is also in pounds per week.

The energy price change is expressed as a percentage.

The sum of these changes for all commodities consumed by the household gives the total cost of the energy price change if the household did not alter its behaviour in response to the change. That is if the household continued to consume the same amount of the commodity regardless of the new price. The adjustment that the household makes to its spending patterns is modelled according to the assumption that if prices increase, this has the same overall effect as an equivalent decrease in income. In these simulations, real income is held constant to isolate the effect the of price changes. The implementation in the model is that following the calculation of the cost due to a change in the price of energy, the new spending pattern is copied from a household of the same demographic type, housing tenure and employment status of the household respondent person (HRP), with an income which is lower than the current household by an amount equal (or as close as possible) to the increase in energy costs. This is repeated until the spending patterns of all households have been adjusted. This process is summarised in the following pseudocode representation.

```
for each year
    for each household
        calculate cost of energy price change (∑ ΔE)

    for each household
        find an existing household of the same demographic type,
        housing tenure and employment status of the HRP that has an
        income similar to the current household's income minus the
```

```
                                                    extra costs
        copy expenditures into temporary household expenditure
                                                    variables

        if no appropriate household exists, inflate or deflate the current
            household's expenditures in proportion to the extra costs

    for each household
        update expenditures from temporary variables
    output results
```

notes: If a household with exactly the same income cannot be found then by a 50/50 chance, a household with the next higher or lower income is chosen. The expenditure variables are not copied directly to the current household because it is possible that this may become the donor household in subsequent copies. Changing household expenditures in this way would cause unwanted side effects so all expenditures are updated simultaneously at the end of each year.

The EFS categories record only consumption related expenditure and do not include money transfers associated with non-consumption items such as income tax, mortgage capital repayments or savings. As these are correlated with household income, the copying procedure causes the total expenditure to appear to vary when there was no change in actual income. To avoid this effect, the budget shares are represented as a proportion of expenditure rather than income. They are calculated as follows. The total spend on each commodity for the whole population is obtained by adding up the amount of expenditure for all households on each commodity. The ‘sum of expenditures’ is the sum of the total expenditure for each commodity. This value is also saved as the ‘original sum of expenditures’ in the first year of the simulation. Next, the budget share for each item is calculated by dividing the total

spending on each commodity by the 'sum of expenditures'. Finally, the total modified spending on each commodity in pounds per week is calculated by multiplying the budget share for each commodity by the 'original sum of expenditures'. The following example describes this process in more detail.

Suppose there are three commodities A, B and C, each with with average expenditures of £40 per week at the start of the simulation. The 'original sum of expenditures' and the 'sum of expenditures' will be £120. Each commodity will take up one third of the total expenditure.

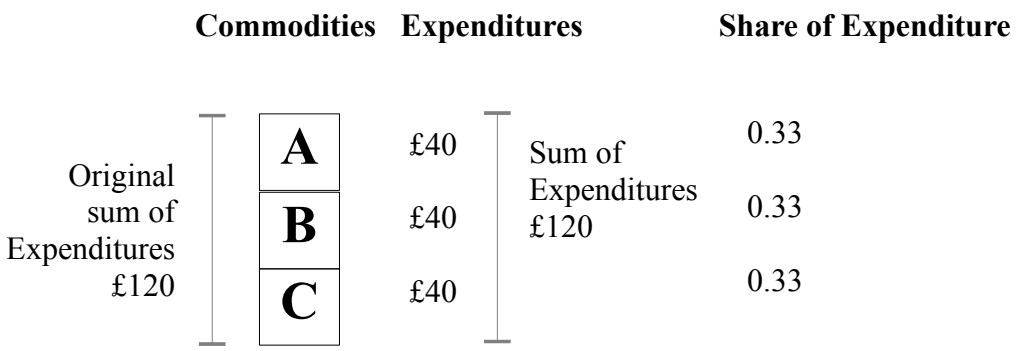


Figure 72: Expenditures and Budget Shares at the Start of the Simulation

After the copying procedure has been run for all households, the expenditure on item C might have risen to £80 per week while the expenditure on items A and B are unchanged.

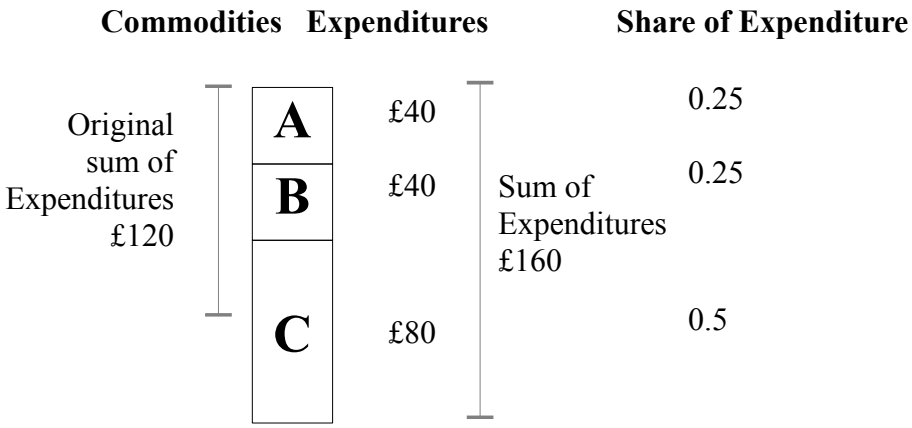


Figure 73: Expenditures and Budget Shares after 1 Year

The 'sum of expenditures' is now £160 per week. This apparent increase in expenditure is a side effect of the copying procedure and in this application is to be cancelled out by rescaling the expenditures. This is done by dividing the new total expenditure for each item by the new 'sum of expenditures'. The budget share for item A is now 40/160 = 0.25. For B it is also 0.25 and for item C it is 80/160 = 0.5.

Now that we have all the budget shares, the expenditures in pounds per week can be calculated by multiplying each by the 'original sum of expenditures'.

Expenditure for item A = 120 * 0.25 = £30
Expenditure for item B = 120 * 0.25 = £30
Expenditure for item C = 120 * 0.50 = £60

Which adds up to £120 as before.

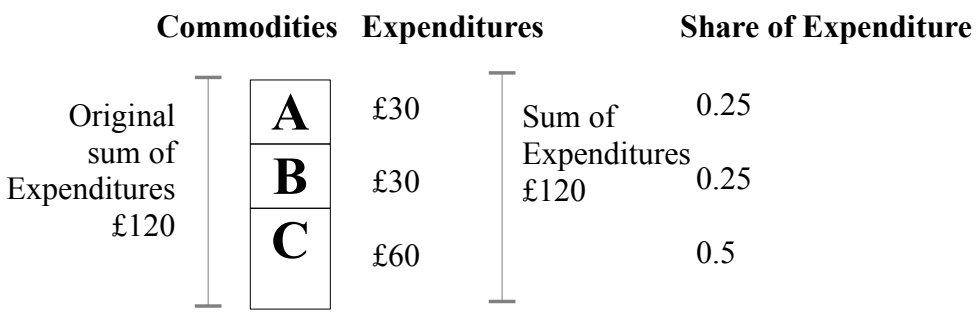


Figure 74: Expenditures and Budget Shares after Rescaling

In practice, the rescaling is not as large as in this example and typically amounts to a few percent.

7.2.2 Results form the Simulations

The first simulation was based on a nominal energy price increase of 10% per year and was run for 20 years. Table 38 shows the expenditure for each commodity in pounds per week at the start of the simulation (year 0) and at the end (year 20). The difference between these is taken as a measure of how spending on a particular item varies as the price of energy changes. The table is sorted into the order of the highest positive change first and the most negative change last.

	Year 0 (£)	Year 20 (£)	Change (£)
Food	46.56	56.30	9.74
Hotels	37.06	39.61	2.56
Other	98.72	100.25	1.53
Housing	46.51	47.98	1.47
Communication	11.55	13.01	1.46
Alcohol	11.02	12.46	1.44
Clothing	22.91	23.53	0.62
Miscellaneous	35.78	35.89	0.11
Education	6.86	6.55	-0.31
Health	5.86	5.46	-0.40
Furniture	30.09	25.01	-5.09
Recreation	58.43	53.21	-5.22
Transport	61.43	53.53	-7.91

Table 38: Change in Expenditures (£ per week)

Table 38 indicates that expenditure on ‘food’ increases the most in response to increases in energy costs. This is followed by ‘hotels’ and ‘other’ but to a lesser extent. ‘Transport’ is the largest faller with ‘recreation’ and ‘furniture’ also decreasing. ‘Communication’ experiences a modest increase of £1.46 per week. The 21 year time series is shown below.

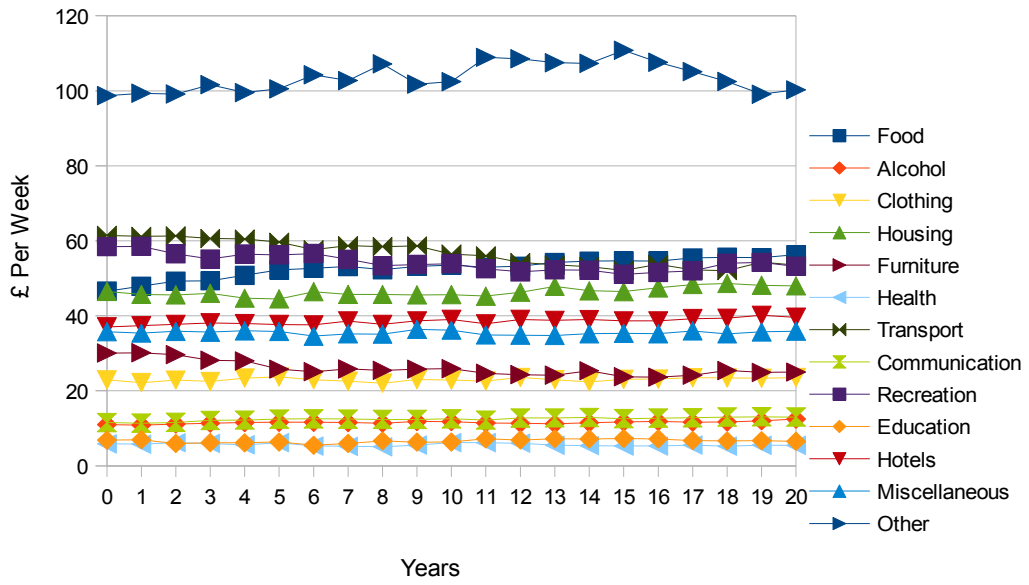


Figure 75: Changes in Expenditure due to a 10% Annual Increase in Energy Prices

It seems surprising that 'other' expenditures, which include mortgage interest payments, shows an increase over much of the series. These costs should not be affected directly by energy prices. They are difficult to change without moving house and it would seem unlikely that households would respond to increased pressure to spend in other areas, by increasing mortgage payments. Further investigation would be required to determine the cause of this anomaly and one possibility is that it is an artefact of the model. It could also be due to some of the many and varied goods within this category. As the interpretation of this spending group is not clear, it might be preferable to remove it from the simulation. One way to do this is firstly by not copying the 'other' expenditures from the donor household when the rest of the expenditures are copied. In addition, 'other' expenditures are excluded from the calculation of the 'sum of expenditures'.

	Year 0 (£)	Year 20 (£)	Change (£)
Food	46.65	55.05	8.40
Hotels	37.26	38.84	1.58
Housing	46.50	47.93	1.43
Alcohol	11.05	12.44	1.39
Communication	11.57	12.44	0.87
Health	5.90	6.43	0.53
Clothing	22.82	23.17	0.35
Other	98.73	98.73	0.00
Miscellaneous	35.93	35.26	-0.67
Education	6.92	6.25	-0.67
Recreation	58.30	56.73	-1.58
Furniture	30.28	27.37	-2.91
Transport	61.17	52.45	-8.72

Table 39: Change in Expenditures when 'Other' Costs are Fixed (£ per week)

In this simulation, there is still an increase in the proportion of expenditure spent on ‘food’ but it is slightly less marked than when ‘other’ costs could vary. ‘Communication’ shows a slight increase. The decrease in ‘Transport’ spending is more pronounced and ‘recreation’ spending declines less than it did when ‘other’ costs could vary.

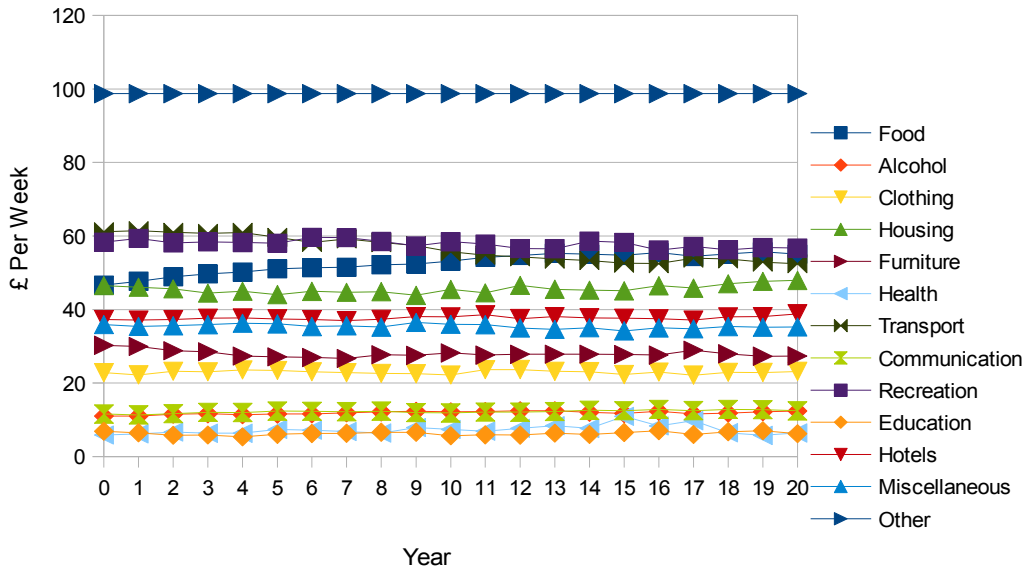


Figure 76: Changes Expenditures when 'Other' Costs are Fixed (£ per week)

These simulations indicate that spending on ‘transport’ would manifest the greatest decrease in response to increases in the price of energy. This seems counter-intuitive because ‘transport’ is one of the spending categories that is most sensitive to energy prices. We might therefore expect ‘transport’ spending to increase when fuel became more expensive. One explanation might be that some individuals may be forced to give up owning a car when the fuel prices become prohibitive. Some households could switch to running one car instead of two and many would decrease the number of miles they drive. All the model tells us is that, as household income falls, the share of expenditure spent on ‘transport’ decreases. Further investigation would be needed to determine any causal link between rising energy prices and changes to the amount spent on ‘transport’. However this result is consistent with a phenomenon known as the 'operating cost effect' where, in the face of rising prices, consumers substitute large items for less expensive goods which use less energy. This is most pronounced

for motor vehicles (Hamilton, 1988).

7.2.3 Discussion

These results appear to support the idea that there is some scope for spending on items like 'transport' and 'recreation' to act as the source of funding for increases in spending on 'communications'. However, these simulations are an example of a novel approach to a complex problem and the conclusions should be treated with caution. Essentially, whether the simulations are an accurate representation of reality depends on a number of assumptions all being correct. These are firstly that a static cross-section of the population as provided by the EFS can be turned into a dynamic model by the assertion that as the financial situation of households changes, they will alter their spending patterns to become more like those who have experienced the new financial situation for some time. The second major assumption underlying this model is that price changes applying to individual commodities can be aggregated together and then thought of as having the same effect as a change in income. Models based on current economic theory would probably find a shift in the budget shares of individual items associated with relative price movements. However, these are set up to isolate the short term or first round response to the price change and do not allow for long term habituation to the new price regime where it could be absorbed into the morass of changes in income, price and social circumstances. The simulations presented in this chapter operate over a time-scale of several years so there may be some plausibility in aggregating price changes. In addition, energy price changes affect nearly all expenditure categories to some extent so amalgamating the variations

becomes more plausible. Despite this, the underlying assertion of treating price changes in the same way as a change in income is at present untested.

7.3 Scenario 2: The Effect of a Shift Towards Homeworking

A second model was developed to explore a scenario where increasing fuel prices leads to a higher prevalence of employees working from home rather than travelling to a central location. Contact with their employer's premises would be facilitated by telecommunications links and this would have implications for the demand placed on the telecommunications network as well as potentially creating opportunities for the development of new products and services.

If a significant proportion of employees exchanged daily commuting for working from home then the money that is currently spent on travelling could become available for telecommunications based services. The question that the model attempts to answer is, how much extra money would households have if there were a large scale shift towards homeworking? Also, how much of this would be consumed in extra household expenses like utility bills inflated by the household being occupied for a greater proportion of the day?

In the past, homeworking often has been associated with low paid work with little control over hours or terms of employment. Moore (2009) contrasts this traditional homeworking with the situation of professional homeworkers who tend to be well paid, highly skilled and have a significant amount of discretion around their working

activities. Another important type of homeworking arrangement is known as teleworking. According to Ruiz & Walling (2005: 417), 'there is no standard definition of teleworking, but it is generally taken to involve working in a location that is separate from a central workplace using telecommunication'. A significant proportion of teleworkers use their home as a base while travelling to a variety of locations where the work is actually carried out. This mode of working has for example, become more common in building related trades. However, this group are unlikely to make significant savings in their travel costs by basing themselves at home because it is still necessary to leave home to reach the work site. For this reason, the study focuses on individuals whose place of work is within the home.

Figure 79 shows the location of work for individuals in the UK aged 16 and over. This was derived from the 2006 BHPS by plotting the variable 'work location' (PJBPL) in the record PINDRESP weighted by the cross sectional respondent weight (PXRWGHT). The bulk of workers (82.5%) travel to their employer's premises. Only 1.86% work at home.

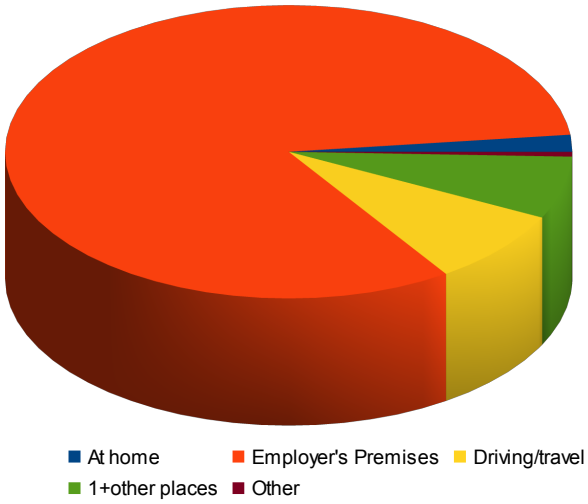


Figure 77: Location of Work in 2006

In a survey of 3000 motorists (elephant.co.uk, 2005), 68% indicated that they would work from home if it were an option - so it seems that there is no shortage of demand for homeworking. This begs the question of why homeworkers currently make up such a low proportion of the working population. Clearly, there are many occupations where it would be impossible or inappropriate to conduct the work at home. However, even where it would be feasible, there are problems to be overcome for both employers and employees. Dwelly & Bennion (2003) find a number of barriers to homeworking for employees. These include a lack of space or suitable work area as well as deficiencies in IT equipment and remote working infrastructure. Other factors include concerns that homeworkers would miss out on promotion opportunities; that colleagues would have to make up for their absence; that working from home might be seen as a lack of commitment to the job and that there could be difficulties in separating work from domestic commitments. For employers, the implementation of homeworking on a large scale can mean quite a significant cultural shift. Practical

issues that would have to be resolved include the provision of IT and office equipment and ensuring that the computer system linking workers, managers and colleagues operates reliably. Health and safety issues would have to be addressed as would any tax and insurance implications. Perhaps the greatest challenge facing employers however, according to Dwelly & Bennion, would be to make the transition from measuring the amount of work done in terms of hours spent, to measuring performance by the quality of the work produced. If this can be done successfully, there are significant rewards for the employers; not least, improvements in productivity.

Despite this, forecast increases in the level of homeworking are quite modest. Felstead (2009) makes some projections based on recent trends using data from the Labour Force Survey (LFS). Although he estimates that 'the proportion using their home as a place of work for one day a week will rise to around a fifth by 2025', the proportion using their home as their main place of work is likely to rise by a 'fraction of a percentage point' (Felstead, 2009: 10).

This model does not attempt to predict what level of homeworking will be achieved in the future. Rather, its aim is to predict what changes in household spending patterns might take place assuming there was a significant increase in homeworking brought about by prohibitively high fuel prices. Nevertheless, it would seem appropriate to define some upper limit that the proportion of homeworkers might approach. Dwelly & Bennion (2003: 4) define homeworking as 'office-type work

carried out by employees working from/at home'. Also, from the ONS (1991), 40% of jobs were located in an office. This can be taken as an upper limit above which homeworking is unlikely to rise without a significant change in the structure of the UK economy.

7.3.1 Data Source

The BHPS includes a set of questions concerning the respondent's work and travel patterns. There are also a small number of household expenditure items as well as a range of demographic variables. This allows a link to be made between travel and expenditure patterns. The initial population for the model consists of a weighted sample of household and individual data. The household file is loaded into the model first. This takes the form of a space delimited text file that was created by selecting variables from the original SPSS file provided by the UK Data Archive.

Table 40 shows some of the household level variables used in the homeworking model as represented in the program.

Variable Name	Description
currenthouse	Household identification number
weight	Household weight to adjust sampled data to represent UK population
housesize	Number of people in the household
xpmg	Annual mortgage payments
renrg	Annual gross amount spent on rent
xpfood	Total annualised food and grocery bill
fihhyr	Annual household income
xplecy	Total annualised electricity bill
xpgasy	Total annualised gas bill
xphsn	Total annual housing costs (sum of rent + mortgage)
xpoily	Total annual amount spent on oil
xpsfly	Total annual amount spent on solid fuel
hsnetb	Presence of broadband within the household

Table 40: Household Level Variables

Once the household file has been read in, individual-level data is obtained from another space delimited text file.

Table 41 shows the most important variables in this file.

Variable Name	Description
case	Household identification number
person	Household member number
nsex	Gender (1 – male: 2 - female)
nstatus	Marital status
age	Current age in years
children	Number of children
nemployment	Employment status
education	Age completed full-time education
ajbhrs	Number of hours worked per week
apaygyr	Gross earnings in the last 12 months
atenure	Housing tenure
worklocation	Place of work (0 – inapplicable: 1 – home: 2-employer's premises: 3 - other
nmeansoftravel	Primary method of travelling to work
traveltime	Minutes spent travelling to work daily

Table 41: Individual Level Variables

Some of the fields are read in as a number and then converted to a string variable in the program for better readability.

nemployment is converted to employment

0 unknown
1 self employed
2 employed
3 unemployed
4 retired
5 family care
6 full-time student

7 long-term sick/disabled
8 other

nmeansoftravel is converted to meansoftravel

0 unknown
1 train
2 underground train
3 bus
4 motorcycle
5 car
6 car passenger
7 bicycle
8 walk
9 other

When both files have been read, individuals are placed in the household designated by their BHPS household identification number. The population is then weighted to correspond to the UK population by adding or deleting households on the basis of their household level weights. In the BHPS, weights are capped at 2.5 to minimise variance inflation and to reduce the incidence of high-value weights (Taylor et al., 2009). This means that there can be at most three copies of any household and its occupants.

The initial level of homeworking can now be calculated. This is done by dividing the number of individuals with a worklocation of 'home' by the number of individuals with a worklocation of either 'home' or 'employer's premises'. This quantity is multiplied by 100 to obtain the percentage of homeworkers within the section of the working population where the place of work is their own home or their employer's premises. The approximately 6% of the whole population whose work location is

'other' remain in this state throughout the simulation.

A target level of homeworking can be set for the following year by moving the slider on the user interface or by entering the number in the set_scenario procedure. If there is a difference between the current percentage of homeworkers and the target level, the number of individuals who would have to change state to reach the target level is calculated. Following this, the required number of individuals are selected at random to either transition from working at their employer's premises to homeworking or transition back to working at their employer's premises as appropriate.

The random selection of individuals for transition implies that all workers have an equal probability of working from home. This is clearly an oversimplification and a more sophisticated selection could be made by taking into account some of the elements that might make homeworking more or less likely for a particular individual. However, this turns out to be a complex task. One approach could be to set transition probabilities according to the current prevalence of homeworking in each industrial sector. Some sectors such as administration have higher current levels of homeworking while others such as mining have low levels. The difficulty with this approach is that we cannot be certain that if there were significant increases in the level of homeworking that these ratios would remain constant. Perhaps, those sectors with currently high levels of homeworking are approaching saturation while there is greatest potential for a rise in homeworking in the sectors where it is currently quite rare. The detailed investigation of the potential for homeworking in each industrial

sector would be quite a major task and lies beyond the scope of the current project. At any rate, the major influence on household spending patterns to be investigated is the location of the work. The difference in consumption patterns between workers in various industrial sectors is assumed to be small in comparison.

Once all the work locations have been set, the next task is to update time spent travelling and mode of transport for each individual who has experienced a transition during the current year. This is done by copying the relevant variables from a suitable donor individual who is already in the new state. The donor must have the same worklocation and employment status as the current individual. They must not have changed worklocation during the current year and must have a similar personal income. In early prototypes of this model it was found that income fell as the level of homeworking increased. This was thought to be due to the presence of traditional homeworkers within the dataset having lower than average incomes. The current scenario is based on the idea of individuals keeping their current job and salary when making the transition to homeworking so this feature was unwanted. Choosing an individual with a similar income effectively controls for income when making transitions in worklocation. When there is a transition, the following variables are copied from the donor: ajbhrs, worklocation, meansoftravel, traveltime, paygyr and employment. They are stored in temporary variables within the recipient agent. When all transitions have been processed, these are copied into the actual agent variables.

When members of a household change their work location, this alters the composition

of the household in terms of the number of homeworkers who live there. It is this state transition that drives variations in household expenditure patterns. For all households where the number of homeworkers has changed during the year, the program searches for another household with a similar annual household income, that also has the same demographic type, number of adults, number of children, number of homeworkers and number who work at their employer's premises. These restrictions arose during discussions with the co-sponsor as characteristics that would not change as a result of a transition to homeworking. If a suitable household is found, its income, monthly mortgage payments, monthly rent payments, gas, electric, oil and solid fuel costs along with an indicator of the presence of broadband internet are copied into temporary household variables. If there is no suitable household, the current household updates its temporary variables from its current variables so in effect nothing happens. After all households have been processed, the temporary variables are copied into the household's working variables.

In the initial dataset, the number of households containing at least one homeworker is about 85, subject to a small random variation due to the weighting procedure. This results in quite a limited pool of donor households, particularly when the restrictions on number of adults, children and household type are applied. As a result, it is quite often the case that a suitable household cannot be found or where there is one, there is a large jump in household income. Typical rates of not finding a donor household are 10% but occasionally this can rise to over 25% in some simulations depending on random variation. This has two effects. One is that there is more variation in the

households which are selected as donors. This can be alleviated somewhat by running the simulation a number of times. The other is that the results from the simulations are likely to underestimate slightly the changes in spending patterns due to increased homeworking.

There is no data on travelling costs, as such, in the BHPS so they are calculated from the amount of time spent travelling to work combined with an estimate of the cost per minute of this journey. The survey of commuters referred to earlier (elephant.co.uk, 2005) found that motorists spend on average £454 a year on fuel, travelling to and from work. The average round trip daily commute was 47 minutes. This implies that every minute of the daily commute, accumulated over the year, costs 454/47 which is £9.66. There is data on time spent commuting in the BHPS and this is multiplied by the cost per minute to find the travelling costs for each individual. The survey also found that 20% of motorists pay to park at work and for this group, the average annual cost amounts to £1,039 (about £20 per week). In the model, this is added to one in five randomly selected commuters who travel by car. The same mileage rates are applied to all modes of transport except walking and cycling. This is partly because the car is the most prevalent mode of transport and partly because for some commuters the modes are substitutes. This would allow for some kind of market to operate where if disproportionate costs applied to one mode of transport, this would become less popular and may eventually lead to decreasing costs.

At the end of each simulated year, the average household expenditures are calculated

for the population in the categories of 'food', 'mortgage', 'rent', 'utilities' and 'commuting'. 'Utilities' is the sum of gas, electric, oil and solid fuel. The method of holding incomes constant was found not to be perfect in that household incomes drift upwards over time. This is thought to be due non-linearities in the income distribution. A 'correctionfactor' was implemented as the previous year's mean household income divided by the current year's mean household income. This is then multiplied by each expenditure category and the current year's income.

7.3.2 Results

Due to the stochastic element within these simulations, it was necessary to run the scenario several times to obtain a consistent result. The following results represent the average values of 10 simulations where the level of homeworking was increased from its 2006 value of slightly over 2% to a nominal 40% of those in employment, in steps of 10% per year. The initial level of homeworkers of 2.1% is slightly higher than the value of 1.86% mentioned earlier. This is due to those with a worklocation of 'other' not being considered for transition.

	Food	**Mortgage**	**Commuting**	**Rent**	**Utilities**
2%	£65.70	£47.36	£15.91	£10.55	£14.85
40%	£65.71	£51.77	£13.75	£10.01	£15.28
Difference	**£0.01**	**£4.42**	**£-2.16**	**£-0.54**	**£0.43**

Table 42: Effect of Widespread Homeworking

Table 42 indicates that if homeworking became widespread, households would on average save £2.16 per week on commuting. This would be offset by a 43p per week

increase in utility bills. Food is virtually unchanged with 1p per week extra being spent. The shift to homeworking is associated with an increase of £4.42 per week on mortgages and a 54p per week decrease in rents. The increase in mortgage spending is greater than the amount saved from not commuting and the decrease in rents. This presumably comes form other items not included in the simulation.

These values are distributed over the whole population and it would also be useful to determine what the expenditure changes are, just for those who experienced a transition to homeworking. The initial prevalence of homeworking was 2.12%. This was increased to 40%; a difference of 37.88%. Increasing the expenditures in linear proportion implies that for each household which experiences a transition from working at their employer's premises: food increases by 3p per week, mortgage spending increases by £11.67, commuting costs decrease by £5.57, rent decreases by £1.43 and utilities increase by £1.14.

It seems clear from these results that a typical household will make some savings from exchanging daily commuting to an employer's premises to working from home. The additional utility costs of £1.14 per week are more than compensated for by savings in commuting of £5.57 per week. This, in conjunction with the inevitable savings in time spent commuting go some way to explaining the high level of interest in working from home that was reported in the elephant.co.uk survey. Also, if energy prices were to increase significantly, it would seem likely that this relationship would continue because both utility costs and travelling costs would both be affected.

7.3.3 Discussion

The second application was to anticipate the effects of a large increase in the proportion of the population working from home. This simulation operated in two stages. In the first, individuals who were currently working at their employer's premises were selected at random to transition to working from home. This resulted in an increase in the number of households where there was at least one homeworker. The second stage was to update the spending patterns of these households, with those from households that already contained one or more homeworkers. The results indicated that the increased utility costs associated with the household being occupied during the day were less than the savings that accrued from abandoning the daily commute to work.

A side effect of this model was that mortgage payments increased while rental spending decreased. This reflects the tendency, in the data, for households containing a homeworker to have a different pattern of housing tenure than households where there are no homeworkers. In some applications, it might be useful to have all correlations related to a particular transition brought into the model. In this case however, it is questionable whether a transition to homeworking would actually cause a change in housing tenure. If so, the model could be amended to control for the level of mortgage and rental payments. Another area where the model could be developed is in the selection of individuals to transition from working at their employer's premises to working from home. The current arrangement is to select at random, however the simulation would assume more realism if a plausible method of selecting

which individuals are more likely to become homeworkers could be found.

7.4 Conclusions and Further Work

This chapter has described two practical applications of models based on a random assignment scheme. The first one attempted to answer the question: if there was a large and sustained rise in the price of energy, what sectors of household spending would be affected? The simulations indicated that spending on 'transport' would be the item that would decrease the most, along with 'recreation' and 'furniture'; while 'food' would be preserved. 'Communication' manifested a small rise despite competing for its share of expenditure with other items. The results seem to support the idea that there is some potential for telecommunications to replace the functions provided by 'transport' and 'recreation'. These conclusions can only be considered as tentative until more confirmation of the validity of the underlying assumptions of the model can be obtained.

As the model contains data for all households, it would be relatively straightforward to represent the results separately for particular sections of the population. This could be in income quartiles, deciles or for some other group of interest. Another potential area for further research would be to vary incomes at the same time as prices. A small percentage change in income might outweigh the effects of the price increases for one section of the population but would be insufficient for another group. This might reveal further differences in how different types of households would be affected. Another area of investigation would be to attempt to consider the effect that

household budget changes would have on future energy prices. It is implicit within the simulations that if household incomes are constant while energy prices rise, there would be a decrease in the total amount of energy used. The reduced demand would exert a downward pressure on prices. If this supply side response was estimated, it could be fed back into the model to determine the next year's price changes.

The homeworking model showed how the random assignment scheme could predict the effects of a discrete change in household conditions. Since incomes are held constant, the variations in expenditure patterns are due largely to behavioural change. It appeared from the results that a move to homeworking would save households a significant amount of money by not travelling to a place of work. This money would then be available to spend on other items such as telecommunications or improved housing. The model is limited in that it does not attempt to predict which individuals are more likely to transition to homeworking.

Despite these limitations, the scenarios presented in this chapter have demonstrated the feasibility of using the random assignment scheme as the basis for a wide range of microsimulation models. The next chapter evaluates the strengths and weaknesses of this approach and identifies what the thesis as a whole adds to the available methods of modelling household expenditure.

Chapter 8: Conclusion

8.1 Introduction

This chapter provides a review the research described in this thesis. It uses the information gained in developing the microsimulation models to answer the research questions posed at the start of the project. Next, it makes an assessment of the significance of the results and what the findings contribute to knowledge of methods and tools for the microsimulation of household expenditure. Finally, it acknowledges the limitations of the research and uses these to identify possible avenues for further work.

In Chapter 1, it was noted that consumer spending represents a substantial component of the UK economy. As a result, the ability to model household spending patterns is of interest to many organisations in both the governmental and commercial sectors. Chapter 2 described several examples were microsimulation models had been applied successfully to model at a disaggregated level. However, it was also suggested that more widespread use of this technology has been hampered by the complexity of the models and a lack of tools for their development. A way to alleviate this problem was proposed in the form of an agent-based modelling toolkit known as NetLogo. Chapter 3 reviewed current methods of economic modelling and argued that the issues of dimensionality, heterogeneity and parametric specification are problematic when applying the usual demand system approaches in the context of microsimulation modelling. Although significant advances are being made, these limitations restrict

the level of disaggregation and number of goods that can be modelled at one time due to data limitations in estimating the parameters for the equations. It was then suggested that random assignment provides a way to avoid these difficulties because there are no parameters to estimate and the method makes it possible to retain the individual cases and their distribution. The potential for NetLogo and random assignment to contribute to modelling household expenditure then formed the subject of investigation for this thesis and was summarised in the question:

> *How can random assignment and NetLogo be combined to develop a coherent micro-level framework for the analysis of household expenditure patterns?*

This main question was divided into two parts:

> *1. How suitable is NetLogo as a platform for developing a dynamic microsimulation model?*
>
> *2. To what extent can random assignment form the basis of an approach for expenditure analysis at the micro-level?*

These general questions were further decomposed into several sub-questions that could be answered objectively from the experience of developing and using the models created during the course of this research. They were:

1. Does NetLogo have sufficient functionality to implement all the usual functions of a dynamic microsimulation model?

2. Does NetLogo have any additional features that are not usually found in existing dynamic microsimulation models?

3. Is its processing speed adequate for the task and how does it compare with an established example?

4. How easy is NetLogo to use: a) from a developer's point of view in creating a new model and b) from an end user's point of view, running a model and obtaining results?

5. Is it feasible to use a random assignment scheme for modelling household spending patterns and are there any difficulties to be overcome?

6. Can a model, implemented using a random assignment scheme, produce standard results that would be expected from previous research?

7. Is random assignment applicable in a wide range of areas and does it have any limitations?

The method of gathering the information needed to answer these questions was to

implement a number of microsimulation models and evaluate the results. Chapter 4 described the development of a dynamic microsimulation model using NetLogo to project the UK population over time. Chapter 5 implemented a model to determine the effects of changing household income using a random assignment scheme. Then, in Chapter 6, the two methods were combined to tackle an important substantive problem which was to model the effects of population ageing on UK household expenditure. Finally, Chapter 7 described two further applications to demonstrate how behavioural change, in the form of responding to price variations and adapting working practices can be implemented using random assignment.

The remainder of this chapter will begin by tackling the list of sub-questions which can be answered with straightforward, objective observations based on the experience of developing and using the microsimulation models. This will provide the basis for answering the two general questions on NetLogo and random assignment. The response to these will then allow the central question to be answered. After this, the significance of the findings will be assessed, firstly in terms of their contribution to methods and tools for microsimulation modelling and then with reference to the challenges for microsimulation modelling that were identified in Chapter 2. Finally, the limitations of the research are discussed and further work is identified.

8.2 Responses to the Specific Questions

1. Does NetLogo have sufficient functionality to implement all the usual functions of a dynamic microsimulation model?

A number of dynamic microsimulation models were reviewed in Chapter 2 and it was found that the processes of household formation that were implemented in a typical model were: mortality, fertility, leaving home, partnership formation and dissolution. The model developed in Chapter 4 included all these processes and so demonstrates that NetLogo, despite its relatively simple scripting language is adequate for this task.

2. Does NetLogo have any additional features that are not usually found in existing dynamic microsimulation models?

Much of the motivation for using an agent-based modelling platform was that the built-in features for manipulating the collection of agents and implementing the graphical user interface (GUI) would reduce the burden of developing the software. In the event, it transpired that these features offered more functionality than is usually found in existing microsimulation models. It can be seen from Figures 22 and 23 that Tyche provides an interface with advanced GUI features such as text boxes for entering model parameters, sliders which can be used to control the parameters while the program is running and charts where the progress of the simulation can be monitored in real time. This results in a much more flexible interface than that of LIAM2 for example, which only has text output or EUROMOD which uses spreadsheets to enter model parameters. In addition, the world grid gives a visual representation of the population and the ‘probes’ allow the state variables of a chosen agent to be monitored as the program runs. NetLogo applications can also be run in the form of an applet which can be accessed remotely over the internet.

3. Is its processing speed adequate for the task and how does it compare with an established example?

The processing speed of Tyche was compared against that of LIAM2 in Chapter 4. It was found that, with a population size of 30,000 households, one year of a dynamic microsimulation including all the usual modules could be completed in 10 seconds. This is likely to be satisfactory in many applications. However, with a dataset of 100,000, cases, Tyche was about seven times slower than LIAM2. Furthermore, execution times in the combined model of Chapter 6 were very long, with one simulated year of a population of 25,000 households taking about 20 minutes to run. This meant that each 30 year simulation took about 11 hours to complete. Some of the processing time will be taken up by the advanced graphical features mentioned above but NetLogo seemed to be particularly slow when the agents had a large number of attributes, as was the case in the income and combined models. If the processing speed of computer technology continues to increase, as it has in recent years, this issue may become less of a problem but currently, NetLogo cannot be recommended for modelling large datasets of several thousand cases where each case has a large number of attributes.

4. How easy is NetLogo to use: a) from a developer's point of view in creating a new model and b) from an end user's point of view, running a model and obtaining results?

As discussed in Chapter 2, a dynamic microsimulation model is a complex piece of software and implementations have been known to cost hundreds of thousands of

pounds to develop. Using NetLogo for this purpose provided an easy to use interface and powerful programming language. The scripting language consists of high-level commands to manipulate the collection of agents but was flexible enough not to restrict the program design. The development environment is packaged into three interfaces – the GUI, the 'info' tab which contains information on how to use the program and the programming tab. Switching between these areas can be done with one click which helps to make the platform easy to use.

From the end user's point of view, NetLogo provides a range of functions that give it an efficient interface. These include buttons to invoke common functions, input boxes and sliders to set parameters, in addition to the world grid and charts to monitor output. All of these items were used in the models described in this thesis and made running the simulations a straightforward exercise.

5. Is it feasible to use a random assignment scheme for modelling household spending patterns and are there any difficulties to be overcome?

The income model described in Chapter 5 and the combined model of Chapter 6 provide evidence of the feasibility of using a random assignment scheme to model household expenditure. One of the main problems to overcome is in model specification. As the matching variables are analogous to the independent variables of a regression equation, the quality of the model depends on how much of the variation in expenditure patterns they capture. Selection of matching variables can be guided by previous research, theory or empirical evidence but as with all approaches, the

choice can be subjective and limited by the available data. Another difficulty is in defining the distance metric or what is meant by a 'similar case'. A variety of functions are available such as Euclidean distance or matching within partitions. It is often not clear at the outset which is the best choice and some experimentation may be useful in guiding the selection. Finally, the stochastic nature of the simulation means that the results vary between runs. This makes it necessary to run the model a number of times to obtain a statistical distribution of the results so that they can be interpreted. Inevitably, this process takes longer to complete and is more cumbersome than if the results could be obtained from one run.

6. Can a model, implemented using a random assignment scheme, produce standard results that would be expected from previous research?

This was investigated in Chapter 5 where the income model showed that overall consumption rises with income but at a slower rate and that budget shares for 'transport' and 'education' are two of the items that increase most with household income, while 'housing' and 'food' exhibit the largest decreases. The last observation is also in agreement with Engel's Law which indicates an inverse correlation between spending on food and income.

7. Is random assignment applicable in a wide range of areas and does it have any limitations?

During the course of this research, random assignment has been applied in a range of applications. In Chapter 5, it was used to model the effect of changes in income on

household expenditure patterns and in Chapter 6, it was used to model demographic change. Then, it was applied, in Chapter 7, to model responses to energy price change and the effects of a rise in homeworking. Furthermore, the method is based on the idea of imputation which is a widely used technique for obtaining values for missing data. As such, it is not restricted to modelling household expenditure. By changing the matching criteria and output variables, it can be adapted to model any relationship within its input data set.

8.3 The General Questions

The responses to the specific questions above, provide the evidence for answering the general questions on NetLogo and random assignment. The first was:

> *1. How suitable is NetLogo as a platform for developing a dynamic microsimulation model?*

The responses, 1 to 4, indicated that NetLogo has the functionality to implement a dynamic microsimulation model while the scripting language reduces the burden of developing the software. NetLogo also has additional features, particularly related to the GUI that give it a high level of usability compared to existing models. However, the processing speed can be slow and potential users should ensure that this will be adequate before developing a model using this platform.

The second general question was:

2. To what extent can random assignment form the basis of an approach for expenditure analysis at the micro-level?

Questions 5 to 7 addressed the feasibility and validity of random assignment. Based on the experience of developing the models described in chapters 5 and 6, it was found that the random assignment approach is feasible and can produce some of the results that have been demonstrated in previous research. As such, it provides the basis for a approach which has the potential to be widely applicable to modelling household expenditure that preserves the individual cases and their distribution.

8.4 The Main Question

Combining the answers to the general questions leads to an answer to the question which provided the central focus for this research:

How can random assignment and NetLogo be combined to develop a coherent micro-level framework for the analysis of household expenditure patterns?

The microsimulation models described in this research demonstrate how NetLogo and random assignment can be used in combination to analyse household expenditure patterns. The claim that the resulting models form a coherent, micro-level framework is founded on the observations firstly that NetLogo is an ABM platform that operates by modelling the individual units and secondly that random assignment operates on individual cases throughout, without the need for aggregation. The combination is

then coherent in its application of micro-level concepts and methods throughout and so forms a framework which has been shown to be applicable, potentially, in a wide range of applications.

8.5 Contribution to Knowledge

While the last section provided answers to the research questions that set the scope of investigation for the thesis, this section evaluates the significance of the findings in terms of what they contribute to knowledge of methods and tools for the microsimulation of household expenditure. This is measured by the extent to which the results have been disseminated in peer reviewed journals, working papers and conference presentations as well as cases where the outputs from the research have been made use of by others.

In essence, the novelty of this thesis arises from applying the agent-based modelling platform, NetLogo, in a new area (microsimulation) and then applying random assignment in applications for which it has not been used before (modelling household expenditure). This provides information on how well each performs in the new areas and so expands the area of applicability of these two technologies. In so doing, it provides a solution to some of the problems that have dogged attempts to model household expenditure at the micro-level using methods and tools that were available previously. The first of these is that the issues identified in Chapter 3 of dimensionality, heterogeneity and specification of functional form have placed a restriction on the number of goods and level of disaggregation that can be represented

in the usual parametric demand system approach. The solution proposed was to apply random assignment and it was shown how this can provide a method of modelling at the micro-level which retains the distribution of cases and does not limit the number of goods. These results were presented at the 3rd General Conference of the International Microsimulation Association (Lawson, 2011) and were later published in the peer reviewed, International Journal of Microsimulation (Lawson, 2013a). The method has also been used in collaborative work to provide the micro-level dataset that was built upon using a QAIDS model and a spatial microsimulation component. This was published in a book chapter (Anderson, De Agostini and Lawson, 2013).

The second problem was the complexity of developing the software for microsimulation modelling. A review of current microsimulation models and their development environments, which formed part of the literature review for this thesis, was published as a TASC (Technology and Social Change) working paper (Lawson, 2008). Options available included using a high-level language such as C++, a preprocessor (Modgen), a database (UMDBS), a framework written in Python interfaced through YAML (LIAM2) and a framework and library ABM (JAS). Using a script based AMB provides another class of approach which was found to provide a powerful and efficient platform for model development. The dynamic microsimulation model Tyche, may be the most accessible of its kind and also provides a framework which is adaptable for new populations by modifying its equation parameters and changing the input data file. The demographic model is described in a CRESI (Centre for Research in Economic Sociology and Innovation)

working paper (Lawson, 2009).

The combination of NetLogo and random assignment was tested on a substantive problem which was to model the effect of population ageing on household expenditure. This demonstrated that the methods can be used to contribute to the debate on this issue, finding that natural demographic change leads to a significant increase in total expenditure, the potential effect of which may be to offset some of the costs of the ageing population. The findings from this part of the research were presented at the 4th General Conference of the International Microsimulation Association (Lawson, 2013b). In addition, two further models were developed, combining NetLogo and random assignment, for use by the co-sponsor of this research, BT. These were to model the effect of changes in the price of energy on household spending patterns and the effect of a rise in the prevalence of working from home, on the way households allocate their budget. The results from this part of the research was used to inform BT's long-term infrastructure planning.

In summary, the contribution that this thesis makes to extend knowledge and methods for microsimulation modelling are condensed into the following list.

1. The advanced graphical features and efficient scripting language mean that NetLogo is probably the easiest to use development platform for dynamic microsimulation modelling and Tyche is the most accessible model.

2. Random assignment makes it possible to model at the micro-level, retaining the heterogeneity of cases without limiting the number of goods.

3. The combination of NetLogo and random assignment provide a coherent micro-level framework for the microsimulation of household expenditure.

4. Population ageing leads to higher expenditure in most categories which may offset some of the costs.

8.6 Impact on Microsimulation

The previous section outlined what contribution the research presented in this thesis makes to the particular area of microsimulation as applied to the modelling of household expenditure. This section considers what the thesis contributes to addressing some of the more general problems associated with microsimulation that were introduced in Chapter 2 which were:

- complexity of software
- cost and development time
- poor usability
- difficult to validate
- limited behavioural representation
- poor accessibility
- lack of predictive power

The observation that Tyche makes it easy for the user to set parameters and view output in charts indicates that the research contributes in the area of 'poor usability'. The finding that NetLogo provides a powerful and easy to use development environment also contributes to user friendliness from a developer's perspective. As a result the research may have a secondary impact on 'cost and development time' because if models are easier to develop, the project may become easier to manage and their cost should decline. The next area of contribution is in 'behavioural representation'. Compared to agent-based modelling, the representation of behaviour in microsimulation modelling has been relatively neglected. Where it is implemented, it is usually represented by an equation derived empirically or based on economic theory. However, the behavioural interpretation of the random assignment procedure is that households are copying the consumption habits of those similar to themselves. As Li and O'Donoghue (2013) note, in microsimulation modelling, the representation of interaction between units is still in its infancy and has been limited to simple copying mechanisms such as Lawson (2011). Hence, while the representation of behaviour applied in this research is not as sophisticated as the complex behaviour found in some ABM implementations, it can be said to lie at the leading edge of behavioural representation in microsimulation. While it is conceptually simple, this is also a powerful technique because whatever behavioural rules were used by the donor household will be copied by the recipient. Furthermore, since this approach is not tied to an equation, it is possible to modify the behaviour by reprogramming the unit. The potential for further work in this area is discussed in the next section. Finally, the innovation that underpins these contributions is the use of NetLogo, where its simple

script based programming language and GUI reduces the complexity of developing a dynamic microsimulation so making this field more accessible to the next generation of microsimulation modellers.

8.7 Further Work

This thesis has applied NetLogo and random assignment in new areas in order to investigate how these two methods can be combined to create a framework for the microsimulation of household expenditure. During the course of this research, a number of avenues for further study arose and this section proposes ways in which they could be followed up.

One of these is the question of how to treat cases for which there is no suitable donor. In the income model of Chapter 5, when incomes were rising, there was no suitable donor for the household that already had the highest income. This was dealt with by simply inflating spending for all categories in proportion to the income rise. However, if there had been some information available on how the extreme cases might behave, such as buying new categories of good for example, then a more flexible behavioural model could be produced. This approach could also be applied to the other cases throughout the population to create a more sophisticated representation of behaviour than has traditionally been employed in microsimulation modelling. The problem here is to determine the behavioural rules for the agents to follow, which is one of the central concerns in agent-based modelling.

Another area for further development derives from the stochastic nature of random assignment. This is an imputation process which, depending on the matching scheme, will give a slightly different result each time it is run. In this research, the aggregated results of a number of simulations were averaged and confidence intervals used to estimate the amount of variation between each run. It would be possible to use multiple imputation to combine the results from a number of simulations, at the level of individual households. This would provide a microdata file, instead of aggregated variables. In the multiply imputed microdata file, the estimators could be calculated using Ruben's methods and results in this form would have more general use because they could be analysed in a variety of ways as if they were an observed dataset. This is feasible if the number of cases is constant but problematic if, as was the case in the demographic model, the population size varies over time.

The last area for further research is matching on a longitudinal sequence instead of a cross-sectional set of observations. If it were possible to combine a number of years of data, for each household and copy from a household that has a similar trajectory over time, this would provide a more accurate picture of the way households respond to changing circumstances. However, as mentioned in Chapter 5, there are no longitudinal surveys that provide detailed information on spending patterns so this research must be left for another time.

8.8 Conclusion

This thesis was motivated by perceived gaps in the available methods and tools for modelling household expenditure. In the case of demand systems, the problem was to model a large number of goods at the micro-level while preserving information on the individual cases as they change over time. In microsimulation, the difficulty was in the complexity of developing the software. Random assignment and NetLogo were applied to develop a coherent micro-level framework for the analysis of household expenditure pattens. The resulting model was found to have the capacity to represent a large number of goods at the micro-level was also relatively easy to develop and use. It was tested on the problem of modelling the economic effects of an ageing population and showed that this phenomenon is likely to increase demand for most expenditure groups, especially housing, health and food.

While these substantive results are quite significant in themselves, the overarching contribution of this thesis is to develop and test the combination of NetLogo and random assignment which now adds to the scope of methods and tools for the microsimulation of household expenditure.

References

Aassve, A., Davia, M., Iacovou, M. and Mazzuco, S. (2005) 'Leaving home and poverty among youth: A cross European analysis', *Working Paper of Institute for Social and Economic Research*, paper 2005-24. Colchester: University of Essex.

Abdulai, A. (2002) 'Household Demand for Food in Switzerland. A Quadratic Almost Ideal Demand System', *Swiss Journal of Economics and Statistics,* 02/2002; 138(I).

Absalon, C. (2009) LATINMOD: A tax-benefit microsimulation model for South America, poster session at the 2nd General Conference of the International Microsimulation Association, Ottawa, Canada, 2009.

Adams, W. J. (2009) *The Life and Times of the Central Limit Theorem*, American Mathematical Society/London Mathematical Society.

Afman, M. R., Chappin, E. J. L., Jager, W. and Dijkema, G. P. J. (2010) Agent-based model of transitions in consumer lighting, In *3rd World Congress on Social Simulation: Scientific Advances in Understanding Societal Processes and Dynamics*, 2010.

Al-Bakour, H. and Markose, S. (2003) Swarm Tutorial, CCFEA, August 2003, Unpublished.

Algers, S., Bernauer, E., Boero, M., Breheret, E., Di Taranto, C., Dougherty, M., Fox, K. and Gabard, J. (1997) Review of Microsimulation Models, Simulation Modelling Applied to Road Transport European Scheme Tests, SMARTEST 1997. http://www.its.leeds.ac.uk/projects/smartest/index.html

Alpay, S. and Koc, A. (2002) 'Household Demand in Turkey: An Application of Almost Ideal Demand System with Spatial Cost Index', Working Paper 0226, Cairo: Economic Research Forum.

Andersson, C. and Nordberg, L. (1998) *A User's Guide to CLAN97: A SAS-program for Computation of Point and Standard Error Estimates in Sample Surveys*, Statistics Sweden.

Ando, A. and Altimari, S. (2004) 'A micro simulation model of demographic development and household's economic behaviour in Italy', *Temi di discussione del Servisio Studi, Banca D'Italia*, Number 533 – December 2004.

Anderson, B., De Agostini, P. and Lawson, T. (2014) Estimating the Small Area Effects Austerity Measures in the UK, in G. Dekkers, M. Keegan and C. O'Donoghue (eds.), *New Pathways in Microsimulation*, Ashgate Publishing Limited.

Andreassen, L. (1993) 'Demographic Forecasting with a Dynamic Stochastic Microsimulation Model', *Discussion Paper no. 85,* Central Bureau of Statistics,

Norway.

Andridge, R. R. and Little, J. A. (2010) 'A Review of Hot Deck Imputation for Survey Non-response', *International Statistical Review*, 78(1), 40-64, April 2010.

Armstrong, J., S. (ed.) (2001) *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Springer.

Aassve, A., Davia, M., Iacovou, M. and Mazzuco, S. (2005) 'Leaving home and poverty among youth: A cross European analysis', Working Paper of Institute for Social and Economic Research, paper 2005-24. Colchester: University of Essex.

Atherton, T., Ben-Akiva, M., McFadden, D. and Train, K. 1990 'Micro-simulation of local residential telephone demand under alternative service options and rate structures' in *Telecommunications Demand Modelling, An Integral View* ed. By A. De Fontenay, M. H. Shugard and D. S. Sibley, Elsevier Publishers B. V.

Atkinson, A. B., Rainwater, L. and and Smeeding, T. M. (1995) 'Income Distribution in OECD Countries', *OECD Social Policy Studies,* No. 18, Paris.

Axelrod, R. (1984) *The Evolution of Cooperation*, Basic Books, ISBN 0-465-02122-0.

Axelrod, R. (1995) A model of the emergence of new political actors, In N. Gilbert and R. Conte (eds.), *Artificial Societies: The Computer Simulation of Social Life*, UCL Press, London.

Bacon, B. and Pennec, S. (2009) Microsimulation, Macrosimulation: Model Validation, linkage and alignment, International Microsimulation Association Conference, Ottawa (Canada), 8-11 June 2009.

Baekgaard, H. and M. Robinson, 1997 'The distributional impact of microeconomic reform in Australia', paper presented at the 5th Nordic Micro-simulation Seminar, Stockholm, June 9-10, 1997.

Baekgaard, H. (2002) 'Micro-macro linkage and the alignment of transition processes: some issues, techniques and examples', National Centre for Social and Economic Modelling (NATSEM) *Technical paper No. 25*.

Baker, P. 1991 'SPEND- The IFS simulation program for energy demand', paper for the IFS Conference Simulation of Tax Reforms, December 1991.

Ballas, D., Clarke, G. P. and Wiemers, E. (2005a) 'Building a Dynamic Spatial Microsimulation Model for Ireland', *Population Place and Space*, 11: 157-172.

Ballas, D., Clarke, G., Dorling, D., Eyre, H., Thomas, B. and Rossiter, D. (2005b)

SimBritain: A spatial microsimulation approach to population dynamics, in *Population, Space and Place* 11, 13-34.

Bank of England (2011) Changes in Bank Rate, Minimum Lending Rate, Minimum Band 1 Dealing Rate, Repo Rate and Official Bank Rate, Bank of England. http://www.bankofengland.co.uk/statistics/rates/baserate.pdf

Banks J., Blundell R. and Lewbel, A. (1997) 'Quadratic Engel Curves and Consumer Demand', *Review of Economics and Statistics* 79, 527-539.

Bao, Q., Kochan, B., Bellemans, T., Janssens, D. and Wets, G. (2013) Investigating micro-simulation error in activity-based travel demand forecasting using confidence intervals, In 92nd TRB Annual Meeting Compendium of Papers DVDtation Research Board, Washington, USA, 2013, Transportation Research Board.

Barigozzi, M., Alessi, L., Capasso, M. and Fagiolo, G. (2009) The Distribution of Households Consumption-expenditure Budget Shares, *Working Paper Series No. 1061 / June 2009*, European Central Bank.

Baroni, E. and Richiardi, M. (2007) 'Orcutt's Vision, 50 years on', *LABORatorio R. Reveli Working Papers Series* 65, LABORatorio R. Revelli, Centre for Employment Studies.

Baroni, E., Eklof, M., Hallberg, H., Lindh, T. and Zamac, J. (2007) Fertility Decisions – Simulation in an Agent-Based Model, in Zaidi, A., Harding, A., and Williamson, P., (2009) *New Frontiers in Microsimulation Modelling*, Ashgate Publishing Limited, UK.

Baroni, E., Zamac, J. and Oberg, G. (2009) 'IFSIM Handbook', *Working Paper 2009: 7*, Institute for Futures Studies, Stockholm 2009.

Barten, A. (1964) 'Consumer Demand Functions under Conditions of Almost Additive Preferences', *Econometrica*, 32, 1-38.

Barten, A. (1969) 'Maximum likelihood estimation of a complete system of demand equations', *European Economic Review*, **1**, 7-73.

Bernardino, J. and Araujo, T. (2009) On positional consumption and technological innovation – an agent-based approach, *WP 04/2010/DE/UECE*, School of Economics and Management, Technical University of Lisbon, ISSN No. 0874-4548.

Berry, S., Levinsohn, J. and Pakes, A. (1995) 'Automobile Prices in Market Equilibrium', *Econometrica*, **63**, 841-90.

Bhrolchain, M. (2005) 'The Age Difference at Marriage in England and Wales: a century of patterns and trends', *Population Trends 120*, National Statistics, London,

HMSO. http://eprints.soton.ac.uk/34801/1/PT120AgeDifference.pdf

Birkin, M. and Malleson, N. (2014) An Investigation of the Sensitivity of a Dynamic Microsimulation Model of Urban Neighbourhood Dynamics, in G. Dekkers, M. Keegan and C. O'Donoghue (eds.), *New Pathways in Microsimulation*, Ashgate Publishing Limited.

Birkin, M. and Clarke, M. (1988) 'SYNTHESIS – A Synthetic Spatial Information System for Urban and Regional Analysis: Methods and Examples', *Environment and Planning* A, 20: 1645-1671.

Blanchet, D. and Minez, S. (2007) 'Projecting Pensions and Age at Retirement in France: Some Lessons from the Destinie I Model', in Zaidi, A., Harding, A., and Williamson, P., (2009) *New Frontiers in Microsimulation Modelling*, Ashgate Publishing Limited, UK.

Blundell, R., Horowitz, J. and Parey, M. (2012) 'Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation', *Quantitative Economics*, 3 (2012), 29-51.

Boelaert, J. (2013) A Neural Network Demand System, CES Working Papers, Documents de Travail di Centre d'Economie de la Sorbonne – 2013.81

Booth, H. (2006) Demographic Forecasting: 1980 to 2005 in review, *International Journal of Forecasting,* 22 (2006) 547 – 581.

Bourdieu, P. (1977) *Outline of a Theory of Practice*, Cambridge and New York, Cambridge University Press.

Bourdieu, P. (1984) *Distinction: a Social Critique of the Judgement of Taste*, Oxton, Routledge.

Bourguignon, F. and Spadaro, A. (2006) 'Microsimulation as a Tool for Evaluating Redistribution Policies', *Working Paper Series*, Society for the Study of Economic Inequality, ECINEQ 2006-20.

Box, G. and Jenkins, G. (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.

Bratman, M. (1987) *Intentions, plans and practical reason*, Harvard University Press, Cambridge, MA.

Bresnahan, T. S., Stern, S. and Trajtenberg, M. (1997) 'Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the late 1980s', *RAND Journal of Economics*, **28**, S17-S44.

Brown, L., Nepal, B., Booth, H., Pennec, S., Anstey, K. and Harding, A. (2011) 'Dynamic Modelling of Ageing and Health: The Dynopta Microsimulation Model', *NATSEM Working Paper 11/14*, NATSEM. http://www.natsem.canberra.edu.au/storage/WP14%20Final%20for%20Web.pdf

Brzezinski, M. (2013) 'Do wealth distributions follow power laws? Evidence from "rich lists"', preprint submitted to *Physica A*, March 2013. http://www.ecineq.org/ecineq_bari13/FILESxBari13/CR2/p98.pdf

Buhmann, B., Rainwater, L., Schmaus, G. and Smeeding, T. (1988) 'Equivalence Scales, Well-Being, Inequality, and Poverty – Sensitivity estimates across 10 countries using the Luxembourg Income Study (LIS) database', *Review of Income and Wealth*, 115-142.

Caldwell, S, and Morrison, R. J. (2000) 'Validation of longitudinal dynamic microsimulation models: experience with CORSIM and DYNACAN' in Mitton L., Sutherland H. and Weeks M. J. (Eds.) *Microsimulation modelling for policy analysis: challenges and innovations,* Cambridge University Press, 200-225.

Carey, J., R. (1999) Life Cycles, Kinship and Growth in Human Populations, Human Development 19, University of California Davis, Winter Quarter, 1999. http://entomology.ucdavis.edu/courses/hde19/lecture7.html

Chattoe-Brown (2009) 'Agent-based modelling and microsimulation: ne'er the twain shall meet?', ESRC/BSPS Seminar Series <http://pcwww.liv.ac.uk/~william/ESRC-BSPS%20MSM%20Seminar%20Series_files/Chattoe-Brown%20edited %20transcript.pdf>

Castiglione, J., Freedman, J. and Bradley, M. (2003) Systematic Investigation of Variability due to Random Simulation Error in an Activity-based Microsimulation Forecasting Model, Transportation Research Record, Vol. 1831, p. 76-88.

Cheesbrough, S. and Scott, A. (2003) 'Simulating Demographic Events in the SAGE Model', London, *ESRC-SAGE Technical Note no. 4*, London School of Economics.

Christensen, L., Jorgenson, D. and Lau, L. (1975) 'Transcendental Logarithmic Utility Functions', *The American Economic Review*, 65, 367-383

Cherchye, L., Crawford, I., De Rock, B. and Vermeulen, F. (2009) Chapter 9 The Revealed Preference Approach to Demand, in Daniel J. Slottle (ed.) *Quantifying Consumer Preferences (Contributions to Economic Analysis, Volume 288)*, Emerald Group Publishing Limited, pp. 247-279.

Chin, S. F. and Harding, A. (2007) 'spatialMSM – NATSEM's Small Area Household Model for Australia', in Gupta, A and Harding, A (eds), *Modelling Our Future: Population Ageing, Health and Aged Care*, Model 22, International Symposia in Economic Theory and Econometrics, Volume 16, Elsevier B. V., Amsterdam, pp. 563-

566.

Chin, S. F., Harding, A., Lloyd, R., McNamara, J., Phillips, B., Vu, Q. *(*2005) *'*Spatial Microsimulation Using Synthetic Small Area Estimates of Income, Tax and Social Security Benefits*'*, *Australasian Journal of Regional Studies*, 11 (3): 303-336.

Christensen, M. (2007) 'Integrability of Demand Accounting for Unobservable Heterogeneity: A Test on Panel Data', *WP14/07*, Institute for Fiscal Studies.

Claro, R., M. and Monteiro, C., A. (2010) 'Family income, food prices, and household purchases of fruits and vegetables in Brazil', *Rev Saude Publica 2010*; 44(6).

Clauss, M. and Schubert, S. (2007) 'Microsimulation Meets General Equilibrium - A New Tool for Applied Policy Analysis', in Zaidi, A., Harding, A. and Williamson, P. (2009) *New Frontiers in Microsimulation Modelling*, Ashgate Publishing Limited, UK.

Clements, M. and Hendry, D. (eds.) (2002) *A companion to Economic Forecasting*, Blackwell Companions to Contemporary Economics, Blackwell Publishers Ltd., Massachusetts USA.

Cohen, P and Levesque, H. (1990) 'Intention is choice with commitment', *Artificial*

Intelligence 42, 3, 213-261.

Collings, D. et al. (1999) 'Agent based consumer modelling', Computing in Economics and Finance, Society for Computational Economics, No. 1352.

Courtioux, P., S. Gregoir and D. Houeto, (2008) The Simulation of the Educational Output over the Life Course: The GAMEO Model, 34.

D'Orazio, M., Di Zio, M. and Scanu, M. (2006) *Statistical Matching: Theory and Practice*, Wiley.

Davidsson, P. (2001) Multi agent based simulation: Beyond social simulation, In S. Moss and P. Davidsson (Eds.), *Multi agent based simulation* (Vol.) 1979), Springer, LNCS series.

Dekkers, G. and Belloni, M. (2009) Micro simulation, pension adequacy and the dynamic model MIDAS: an introduction.

De Agostini, P. and Anderson, B. (2005) 'WMM D3.3 'Utilities trends, elasticities and cohort analyses', Ipswich: Chimera, University of Essex.

Deaton, A. (1988) 'Quality, Quantity, and Spatial Variation of Price', *American Economic Review* (78): 418-30.

Deaton, A. and Muellbauer, J. (1980a), 'An Almost Ideal Demand System', *The American Economic Review*, Vol. 70, No 3, pp. 312-336.

Deaton, A. and Muellbauer, J. (1980b), *Economics and Consumer Behavior*, Cambridge, England: Cambridge University Press.

Dekkers, G. (2011) 'Liam2 A sneak preview', Paper presented at the Treasury Brown Bag, February 2011.

Dijkstra, E. W. and Dahl, O. J. et al. (1972) *Structured Programming*, London, Academic Press, (Chs. 23, 32).

Dilnot, A., Warner, N. and Williams, J. (2011) Fairer Care Funding, The Report of the Commission on Funding of Care and Support, July 2011.

Dittmar, M. (2009) The Future of Nuclear Energy: Facts and Fiction, The Oil Drum: Europe http://europe.theoildrum.com/node/5631#more

Doggett & Ayesha (2009) Global energy demand seen up by 44 percent by 2030, Reuters, Washington http://www.reuters.com/article/idUSN2719528620090527

Duncan, A. (1991) 'A Microsimulation model of labour supply for UK tax reform',

University of Konstanz discussion paper, Series II No. 153, Konstanz.

Dwelly, T. & Bennion, Y. (2003) Time to go home: Embracing the homeworking revolution, The Work Foundation, April 2003 http://www.ictknowledgebase.org.uk/fileadmin/ICT/pdf/Home_sweet_home/Time_to_go_home-1.pdf

Economic Research Service (2009) *Commodity and Food Elasticities: Glossary*, US Department of Agriculture. http://www.ers.usda.gov/Data/Elasticities/glossary.htm

Edmonds, B. and Chattoe-Brown, E. (2005) When Simple Measures Fail: Characterising Social Networks Using Simulation, Presented at the Social Network Analysis: Advances in Empirical Applications Forum, Oxford, July 16-17 2005, CPM Report 05-158, MNU.

Edmonds, B. and Meyer, R. (Eds.) (2013) *Simulating Social Complexity: A Handbook*, Springer.

Elephant.co.uk, (2005) Daily commute is 47 minutes of hell for uk motorists, Press Release http://www.elephant.co.uk/contactUs/pressOffice/pressReleases/daily-commute-is-47-minutes-of-hell-for-UK-motorists.php

Eliasson, G. (1988) “MOSES” - A Presentation of the Swedish Micro-to-Macro

Econometric Model, Preliminary paper.

Emmerson, C., Reed, H. and Shephard, A. (2004) An Assessment of Pensim2, Institute for Fiscal Studies, WP04/21. http://eprints.ucl.ac.uk/2893/1/2893.pdf

Engel, E. (1857) Die Produktions und Consumtionsverhaltnisse des Kaonigreichs Sachsen, reprinted with Engel (1895), Anlage 1, 1-54.

Engel, E. (1895), Das Lebenskosten belgischer Arbeiterfamilien frÄuher und jetzt, *Bulletin de Institut International de Statistique* 9: 1-124.

Epstein, J. M. and Axtell, R. (1996) *Growing Artificial Societies: Social Science from the Bottom Up*, Brookings Institution Press.

Ettlinger, M. and O'Hare J (1996) 'Revenue and Incidence Analysis of State and Local Tax Systems: Tools, Technique and Tradition', Paper presented at the National Tax Association's 89th Annual Conference, 1996.

EUROSTAT (1996) Comparison in real terms of the aggregates of ESA, results for 1994.

Favreault, M. and Smith, K. E., (2004) A primer on the Dynamic Simulation of Income Model (DYNASIM3) Urban Institute.

http://www.urban.org/publications/410961.html

Felstead., A. (2009) Detaching work from place: charting the progress of change and its implications for learning, Beyond Current Horizons, January 2009 http://www.beyondcurrenthorizons.org.uk/wp-content/uploads/final_felstead_detachingworkfromplace200901162.pdf

Flood, L., Jansson, F., Pettersson, T., Sundberg, O. and Westerberg, A. (2005) SESIM III – a Swedish dynamic microsimulation model, Handbook of SESIM, Ministry of Finance, Stockholm.

Forrester, . W. (1961) *Industrial Dynamics*, Pegasus Communications.

Fredriksen, D. (1998) 'Projections of Population, Education, Labour Supply and Public Pension Benefits: Analysis with the Dynamic Microsimulation Model MOSART'. *Social and Economic Studies*, Statistics Norway, Oslo.

Fredriksen, D. and Stølen, N. (2007) 'MOSART', in Gupta, A. and Harding, A. (eds), *Modelling our Future: Population, Ageing, Health and Aged Care*, North-Holland, Amsterdam.

Gardner, M. (1970) 'Mathematical Games: The fantastic combinations of John Conway's new solitaire game "Life"', *Scientific American* 223: 120–123.

Gilbert, N. and Troitzsch, K. (2005) *'Simulation for the Social Scientist – second edition*', England, Open University Press.

Giles, C. and McCrae, J. (1995) TAXBEN: the IFS microsimulation tax and benefit model, *IFS Working Paper, 95/19*, Institute for Fiscal Studies.

Gintis, H. (2000) 'Beyond *Homo economicus*: evidence from experimental economics', *Ecological Economics*, Volume 35, Issue 3, December 2000, Pages 311 – 322.

Gipps, P.G. (1981) 'A behavioural car-following model for computer simulation', Transportation Research Board, 15-B, pp. 105-111.

Goedeme, T., Bosck, K., Salanauskaite, L. and Verbist, G. (2013) 'Testing the Significance of Microsimulation Results: A Plea', *International Journal of Microsimulation*, (2013) 6(3) 50-57.

Goldberger, A., (1989) 'Economic and mechanical models of intergenerational transmission', *American Economic Review*, pp 504-513.

Goldstein, J. (1999) 'Emergence as a Construct: History and Issues' *Emergence: Complexity and Organisation* 1(1): 49-72.

Goran, J. (2001) Activity based travel demand modelling – a literature study, Technical report, Danmarks Transport Forskning.

Hamilton, J. D. (1988) 'A neoclassical model of unemployment and the business cycle', *Journal of Political Economy*, 96 (1988), pp. 593-617.

Hancock, R. Mallender, J. and Pudney, S. (1992) 'Constructing a Computer Model for Simulating the Future Distribution of Pensioner's Incomes for Great Britain' in Hancock and Sutherland (eds.) *Microsimulation Models for Public Policy Analysis: New Frontiers*, STICERD Occasional Paper no. 17, London School of Economics, London.

Hancock, R., Juarez-Garcia, et al. (2006) 'Winners and Losers: Assessing the Distributional Effects of Long-term Care Funding Regimes', ISER Working Paper 2006-43, Colchester, University of Essex.

Hancock, R., Wittenberg, R., Pickard, L., Comas-Herrera, A. and Juarez-Garcia, A. (2007) Paying for Long-Term Care for Older People in the UK: Modelling the Costs and Distributional Effects of a Range of Options, PSSRU Discussion Paper 2336/2, April 2007. http://www.pssru.ac.uk/pdf/dp2336_2.pdf

Harding, A. (2007) 'Challenges and Opportunities of Dynamic Microsimulation

Modelling', *Plenary paper presented to the 1st General Conference of the International Microsimulation Association*, Vienna, 21 August 2007.

Harding, A., Keegan, M. and Kelly, S. (2010) 'Validating a dynamic microsimulation model: Recent experience in Australia', *International Journal of Microsimulation,* (2010) 3(2) 46-64.

Hartmann, S. (1996) 'The World as a Process. Simulations in the Natural and Social Sciences', in Hegselmann et al. (1996) *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, Theory and Decision Library, Dordrecht: Kluwer.

Harvey, C. (2012) Random Walk with Drift, The Free Dictionary. http://financial-dictionary.thefreedictionary.com/Random+walk+with+drift

Hassan, S., Pavon, P., and Gilbert, N. (2008) 'Injecting Data into Simulation: Can Agent-Based Modelling Learn from Microsimulation?' *Proceedings of the World Congress on Social Simulation*, Fairefax, United States.

Hausman, J. A. (1996) 'Valuation of New Goods under Perfect and Imperfect Competition' in T. F. Bresnahan and R. Gordon (eds), *The Economics of New Goods*, Chicago: National Bureau of Economic Research.

Hausman, J. A., Leonard, G. and Zona, J. D. (1994) 'Competitive Analysis with Differentiated Products', *Annales D' Economie at de statistique*, **34**, 159-80.

Hayes, B. (2009) 'Everything is Under Control', *The American Scientist*, Volume 97, May-June 2009.

Henrickson, L. (2002) 'Old Wine in a New Wineskin: College Choice, College Access Using Agent-Based Modelling', *Social Science Computer Review*, Winter 2002 Vol. 20 no. 4 400-419.

Heppenstall, A. J., Crooks, A. T., See, L. M. and Batty, M. (Eds.) (2012) *Agent-based Models of Geographical Systems*, Dordrecht, Springer.

Heppenstall, A., Evans, A. and Birkin, M. (2006) 'Using hybrid agent-based systems to model spatially-influenced retail markets', *Journal of Artificial Societies and Social Simulation*, 9(3), 2.

Hills, J. (2006) 'From Beveridge to Turner: demography, distribution and the future of pensions in the UK', *Journal of the Royal Statistical Society*: series a, 169 (4). pp. 663-679.

Hills, S., Thomas, R. and Dimsdale, N. (2010) The UK Recession in Context – what do three centuries of data tell us?, in Quarterly Bulletin 2010 Q4, Vol. 50, No. 4. Bank

of England.
http://www.bankofengland.co.uk/publications/Documents/quarterlybulletin/threecenturiesofdata.xls

Holm, E., Mäkilä, K. and Lundevaller, E. (2009) 'Imitation or Interaction in Spatial Micro Simulation', Presentation at the 2nd General Conference of the International Microsimulation Association, Ontario, Canada.
http://www.statcan.gc.ca/conferences/ima-aim2009/session3i-eng.htm#a44

Houthakker, H. (1950) 'Revealed preference and the utility function', *Econometrica*, **17**, 159-174.

Hubbert, M. K. (1956) 'Nuclear Energy and the Fossil Fuels', *Drilling and Production Practice*, American Petroleum Institute.
http://www.hubberpeak.com/Hubbert/1956/1956.pdf

Hussenius, J. and Selén, J. (1994) Skatter och socialförsäkringar över livscykeln - En simuleringsmodell, Rapport till ESO, Ministers of Finance DS 1994:135, Stockholm.

Iltanen, S. (2012) Cellular Automata in Urban Spatial Modelling, In A. J. Heppenstall et al., (Eds.) *Agent-Based Models of Geographical Systems*, Springer.

Imhoff, E., and Post, W. (1998) Microsimulation methods for population projection,

in Population, 10e annee, n 1, 1998, pp. 97-136.

Inagaki, S. (2010) Overview of INAHSIM: A Microsimulation Model for Japan, PIE/CIS Discussion Paper 468.

Ingram, D. D., O'Hare, J., Scheuren, F. and Turek, J. (2001), 'Statistical Matching: a New Validation Case Study'. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 746-751.

INSEE (Institut National de la Statistique et des Etudes Economiques) (1999) Le Modele de Microsimulation Dynamique, DESTINIE. Document de travail, G9913. Paris: INSEE.

Intriligator, M. D. (1978) *Econometric Models Techniques, and Applications*, North-Holland Publishing Company, Amsterdam, Oxford.

Jing, M., Heppenstall, A., Harland, K. and Mitchell G. (2014) 'Synthesising carbon emission for mega-cities: A static spatial microsimulation of transport CO2 from urban travel in Beijing', *Computers, Environment and Urban Systems, Volume 45*, May 2014, Pages 78-88, ISSN 0198-9715.

Kahneman, D., Thaler, R., H. (2006) 'Anomalies: Utility Maximization and Experienced Utility', *Journal of Economic Perspectives*, 20(1): 221-234.

Kalb, G. and Thoresen, T. (2007) 'Behavioural Microsimulation: Labour Supply and Child Care Use Responses in Australia and Norway' in Zaidi, A., Harding, A., and Williamson, P., (2009) *New Frontiers in Microsimulation Modelling*, Ashgate Publishing Limited, UK.

Kalton, G. and Kaspyzyk, D. (1986), 'The Treatment of Missing Survey Data', *Survey Methodology*, 12, 1–16.

Keegan, M. and Kelly, S. (2009) 'APPSIM – Dynamic microsimulation modelling of social security and taxation', *Working Paper 14*, August 2009. NATSEM.http://www.natsem.canberra.edu.au/storage/WP14-Social%20security %20and%20taxation-final.pdf

Kheir, N. A. (1988) *Systems Modelling and Computer Simulation,* Electrical Engineering and Electronics, Vol. 1, Marcel Dekker, New York, NY.

Khor, K., Chavis, N., Lovett, S. and White, D. (1995) *IBM Smalltalk Tutorial*, IBM. http://www.inf.ufsc.br/poo/smalltalk/ibm/tutorial/oop.html

Klevmarken, A., Andersson, I., Brose, P., Flood, L., Olovsson, P. and Tasiran, A. (1992) 'MICROHUS. A Micro-simulation Model for the Swedish Household Sector. A Progress Report'. Paper presented at the International Symposium on Economic

Modelling, August 18-20, Gothenburg, Sweden.

Klevmarken, N. A. (1997) 'Behavioural modelling in micro simulation models. A Survey', *Papers 1997-31*, Uppsala – Working Paper Series.

Klevmarken, A. (2008) Chapter 2 Dynamic Microsimulation for Policy Analysis: Problems and Solutions, in Anders Klevmarken, Bjorn Lindgren (ed.) *Simulating an Ageing Population: A Microsimulation Approach Applied to Sweden (Contributions to Economic Analysis, Volume 285*), Emerald Group Publishing Limited, pp.31-53.

Klevmarken, A. and Olovsson, P. (1996) 'Direct and behavioural effects of income tax changes - simulations with the Swedish model MICROHUS', in A. Harding (ed.), *Micro-simulation and Public Policy*, Elsevier Science Publishers, Amsterdam.

Kuhn, T., S. (1962) *The Structure of Scientific Revolutions*, 1st. ed., Chicago Pr.

Kuhn, T. S. (1969) Postscript, in *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press: 174-210.

Lancaster, J. (1971) *Consumer Demand: A New Approach*, New York: Columbia University Press.

Lawson, T. (2008) 'Methods and Tools for the Microsimulation and Forecasting of

Household Expenditure - A Review', *Technology and Social Change Working Papers*, 2008/01.
http://repository.essex.ac.uk/2307/1/TWP-2008-01-Budget-Microsim-03-09-08.pdf

Lawson, T. (2009) 'A Demographic Microsimulation Model Using NetLogo', Working Paper. Centre for Research in Economic Sociology and Innovation *(CRESI) Working Paper 2009-02*, University of Essex, Colchester, UK.
http://repository.essex.ac.uk/2302/1/CWP-2009-02-Microsim-Netlogo.pdf

Lawson, T. (2011) An Agent-based Model of Household Spending Using a Random Assignment Scheme, Paper presented at the 3rd General Conference of the International Microsimulation Association, Stockholm, 8-10 June

Lawson, T. (2013a) 'Modelling Household Spending Using a Random Assignment Scheme', *International Journal of Microsimulation, Vol. 6(2)*, Autumn 2013.
http://www.microsimulation.org/IJM/V6_2/3_IJM_6_2_2013_Lawson.pdf

Lawson, T. (2013b) The Effect of Population Ageing on Aggregate UK Household Expenditure, Paper presented at the 4rd General Conference of the International Microsimulation Association, Canberra, 11-13 December

Lefebvre, M. (2006) 'Population ageing and consumption in Belgium', *CREPP Working Papers 0604*, Centre de Recherche en Economie Publique et de la

Population (CREPP) (Research Center on Public and Population Economics) HEC-Management School, University of Liège.
http://ideas.repec.org/p/epp/wpaper/0604.html

Leombruni, R., and Richiardi, M. (2005) 'LABORsim: an Agent-based Microsimulation of Labour Supply. An Application to Italy', *LABORatorio R. Revelli Working Papers Series 46*, LABORatorio R. Revelli, Centre for Employment Studies.

Lewbel, A. and Pendakur, K. (2009) 'Tricks with Hicks: The EASI Demand System', *American Economic Review*, 99(3): 827-63.

Lewbel, A. and Pendakur, K. (2013) 'Unobserved Preference Heterogeneity in Demand Using Generalised Random Coefficients', Mimeo, Boston College.

Li, J. and O'Donoghue, C. (2013) 'A survey of dynamic microsimulation models: uses, model structure and methodology', *International Journal of Microsimulation*, autumn 2013, 6(2) 3-55, International Microsimulation Association.

Li, J. and O'Donoghue, C. (2014) 'Evaluating Binary Alignment Methods in Microsimulation Models', *Journal of Artificial Societies and Social Simulation*, 17 (1) 15.

Lianos, I. and Genakos, C. (2013) Econometric evidence in EU competition law: an

empirical and theoretical analysis, in I. Lianos and D. Geradin (eds.) *Handbook on European Competition Law: Enforcement and procedure*, Edward Elgar Publishing Cheltenham, UK.

Liegeois, P. and Dekkers, G. (2011) ‘Combining EUROMOD and LIAM tools for the development of dynamic cross-sectional microsimulation models: a sneak preview’, *3rd General Conference of the International Microsimulation Association*, June 8th to 10th, Stockholm, Sweden.

Lietz, C. and Mantovani, D. (2006) 'Lessons from Building and Using' EUROMOD, *EUROMOD Working Paper*, EM5/06.

Loeb, B. S. (1955) 'The Use of Engel's Laws for Predicting Consumer Expenditures', *The Journal of Marketing*, Vol. 20, No. 1 Jul. 1955, pp. 20-27, American Marketing Association, http://www.jstor.org/stable/1248157

Luhrmann, M. (2008) Effects of population ageing on UK consumer demand, IFS CEMMAP, London. http://www.econ.upf.edu/docs/seminars/luhrman.pdf

Lutero, G. (2010) 'The aggregation problem in its historical perspective: a summary overview', Methods development of quarterly national accounts. ISTAT, National Accounts Directorate.

Lutz, W., Goldstein, J. R. and Prinz, C. (1996) Alternative approaches to population projection. In W. Lutz (Ed.), *The future population of the world. What can we assume today?* (pp. 14 – 44), London: Earthscan.

Lutz, W., Sanderson, W., Scherbov, S. and Goujon, A. (1996) World population scenarios for the 21st century. In W. Lutz (Ed.) *The future population of the world: What can we assume today?* (pp. 361 – 396), London: Earthscan.

Lymer, S., Brown, L., Yap, M. and Harding, A. (2008) 'Regional Disability Estimates for New South Wales in 2001 Using Spatial Microsimulation', *Applied Spatial Analysis and Policy,* 1 (2): 99-116.

Mahdavi, B., O'Sullivan, D., and Davis, P. (2007) 'An Agent-based Microsimulation Framework For Investigating Residential Segregation Using Census Data', In Oxley, L. and Kulasiri, D. (eds) *MODSIM07 International Congress on Modelling and Simulation* pp. 365-371.

Malleson, N. (2012) Using agent-based models to simulate crime, In A. J. Heppenstall, A. T. Crooks, L. M. See and M. Batty (Eds.), *Agent-based models of geographical systems* (pp. 411-434), Dordrecht: Springer.

Maria, A. (1997) 'Introduction to Modelling and Simulation', Proceedings of the 1997 Winter Simulation Conference ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B.

L. Nelson.

Maynard, H. H., and Beckman T. N. (1946) *Principles of Marketing,* 4th ed. The Ronald Press Company, New York.

McAleer, M., Medeiros, M. and Slottje, D. (2008) 'A neural network demand system with heteroskedastic errors', *Journal of Econometrics*, 147 (2008) 359-371.

McFadden, D. (1973) 'Conditional Logit Analysis of Qualitative Choice Behaviour' in P. Zarembka (ed.), *Frontiers of Econometrics*, New York: Academic Press.

McLoughlin, F., Duffy, A. and Conlon, M. (2013) 'Evaluation of time series techniques to characterise domestic electricity demand', *Energy*, Volume 50, 1, February 2013, Pages 120-130, Elsevier.

Merz, J. (1993) 'Microsimulation as an Instrument to Evaluate Economic and Social Programmes', *Discussion Paper No. 5,* Social Science Research Network. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1417318

Mikula, B. D., Elias, F., Holmgren, N., Lundkvist, L., (2003) The Swedish Pension Model in context of the pension reform, RFV, Mimesis group 25.

Mitton, L., and Sutherland, H. (1999) Maintaining and Updating POLIMOD,

Docstore, http://www.docstoc.com/docs/70798417/Maintaining-and-Updating-POLIMOD

Mohtashami, T. and Salami, H. (2009) Assessing Predictive Performance of Alternative Time-Series Methods in Forecasting Dairy Products Consumption, International Conference on Applied Economics, ICOAE 2009.

Montgomery, D. C., Jennings, C. L. and Kulahci, M. (2008) *Introduction to Time Series Analysis and Forecasting*, Wiley Series in Probability and Statistics, John Wiley & Sons Inc., USA.

Montgomery, D. C., Peck, E. A. and Vining, G. (2001) *Introduction to Linear Regression Analysis, 3rd Edition,* Wiley Series in Probability and Statistics, John Wiley & Sons Inc., USA.

Moore, J. (2009) 'Making Home Work: A Qualitative Study of Homeworking in the North of England and Wales'. Presentation at the Histories of the Home Subject Specialist Network Symposium, Manchester.

Morand, E., L. Toulemon, S. Pennec, R. Baggio, and F. Billari, (2010) Demographic modelling: the state of the art, *SustainCity Working Paper*, 2.1a, Ined, Paris.

Morrison, R. (2006) Make it so: Event alignment in dynamic microsimulation,

DYNACAN paper.

Morrison, R. (2008) ‘Validation of Longitudinal Microsimulation Models: DYNACAN Practices and Plans’, *Working Paper No. 8,* March 2008, NATSEM.

Moschini, G. (1998) 'The semiflexible almost ideal demand system', *European Economic Review*, Volume 42, Issue 2, Pages 349-364.

Moschini, G. and Morrow, D. (1996) ‘Structural Change and Demand Analysis: a cursory review’, *European Review of Agricultural Economics*, (1996) 23(3): 239-261.

Mullan, P. (2000) *The Imaginary TimeBomb: Why An Ageing Population Is Not A Social Problem,* London, I B Taurus & Co. Ltd.

NetLogo (2010) *NetLogo User Manual*, Version 4.1.1., August 4 2010. http://ccl.northwestern.edu/netlogo/faq.html

Neufeld, C. (2000) Alignment and variance reduction in DYNACAN, In A. Gupta & V. Kapur (Eds.) Microsimulation in Government Policy and Forecasting (pp. 361-382), Amsterdam: North-Holland.

von Neumann J (1966) in Burks A: *The Theory of Self-reproducing Automata*, Urbana, IL: Univ. of Illinois Press.

Nevo, A. (2001) 'Measuring Market Power in the Ready-to-eat Cereal Industry', *Econometrica*, **69**, 307-42.

Nikolai and Madey (2009) 'Tools of the Trade: A Survey of Various Agent Based Modelling Platforms'. *Journal of Artificial Societies and Social Simulation* 12(2)2 <http://jasss.soc.surrey.ac.uk/12/2/2.html>.

NIST/SEMATECH (2003) e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc444.htm#AR

OECD (2012) 'What are Equivalence Classes', OECD Project on Income Distribution and Poverty, http://www.oecd.org/els/soc/OECD-NoteEquivalenceScales.pdf

O'Donoghue, C. (2001a) 'Dynamic Microsimulation: A Methodological Survey', *Brazilian Electronic Journal of Economics,* Department of Economics, Universidade Federal de Pernambuco, Vol. 4(2), December.

O'Donoghue, C. (2001b) Redistribution in the Irish Tax-benefit System, Unpublished Ph.D. Dissertation, London School of Economics.

O'Donoghue, C. (2001c) 'Redistribution over the Lifetime in the Irish Tax-Benefit

System: An application of a Prototype Dynamic Microsimulation Model for Ireland', *The Economic and Social Review*, Vol. 32, No. 3, October 2001, pp. 191-216.

O'Donoghue, C., Baldini, M. and Mantovani, D. (2004) 'Modelling the Redistributive Impact of Indirect Taxes in Europe: An Application of EUROMOD', *EUROMOD Working Paper* No. EM7/01.

O'Donoghue, C., Lennon, J., and Hynes, S. (2009) 'The Life-Cycle Income Analysis Model (LIAM): A Study of a Flexible Dynamic Microsimulation Modelling Computing Framework', I*nternational Journal of Microsimulation*, International Microsimulation Association, vol. 2(1), pages 16-31.

O'Hare, J. F. (2000) 'Impute or Match? Strategies for Microsimulation Modelling', in Gupta, A. and Kapur, V. (eds) 2000 *Microsimulation in Government Policy and Forecasting*, Elsevier, London.

One Asia Capital (2014) Technical Analysis.
http://www.oneasiacapital.com/ProductsServices/TechnicalAnalysis

ONS (1996) 'Birth Statistics, Review of the Registrar General on births and patterns of family building in England and Wales', 1994, London: The Stationary Office.

ONS (2007a) 'National Population Projections', 2006 based projections, ONS October 2007.

ONS (2007b) Household Final Consumption Expenditure Summary 0.CN, *Consumer Trends – Q1 2007*, Office for National Statistics. http://www.ons.gov.uk/ons/rel/consumer-trends/consumer-trends/q1-2007/index.html

ONS (2008) Family Spending: 2007 Edition, Ed Dunn (Ed.), Palgrave MacMillan. http://www.ons.gov.uk/ons/rel/family-spending/family-spending/2007-edition/index.html

ONS (2010a) Household Final Consumption Expenditure Summary 0GS.CN, *Consumer Trends – Q2 2010*, Office for National Statistics. http://www.ons.gov.uk/ons/rel/consumer-trends/consumer-trends/q2-2010/index.html

ONS (2010b) Live Births, Total Fertility Rates in England and Wales, http://www.statistics.gov.uk/cci/nugget.asp?id=369

ONS (2011a) 2010-based Period and Cohort Life Expectancy Tables, ONS October 2011. http://www.ons.gov.uk/ons/rel/lifetables/period-and-cohort-life-expectancy-tables/2010-based/p-and-c-le.html

ONS (2011b) National Population Projections, 2010-based reference volume: Chapter 3: Fertility, Office for National Statistics, chdtfrandcfs61_tcm77-258103[1].xls, http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2010-based-

reference-volume—series-pp2/fertility.html#tab-Fertility-assumptions-for-the-UK.

ONS (2012) Information Paper, Quality and Methodology, Office for National Statistics, December 2012.

ONS (2013a) Historic Interim Life Tables, United Kingdom 2004-2006, Office for National Statistics, iltuk0810reg_tcm-231999[1].xls

ONS (2013b) National Population Projections, 2012-based Statistical Bulletin, Office for National Statistics, November 2013.

Orcutt, G. (1957) 'A new type of socio-economic system', *Review of Economics and Statistics,* 39(2), 116-123, Reprinted in the *International Journal of Microsimulation,* 2006 1(1), 2-8.

Orcutt G. H., Greenberger M., Korbel J. et al. (1961) *Microanalysis of Socioeconomic Systems: a simulation study*, New York: Harper and Row

Orcutt G. H., Caldwell S. and WertheimerII R (1976) *Policy Exploration through Microanalytic Simulation*, Washington D.C.: The Urban Institute.

O'Sullivan, A. (2003) *Economics: Principles in Action*, Upper Saddle River, New Jersey 07458: Pearson Prentice Hall.

Papoulis, A. (1984) *Probability, Random Variables and Stochastic Processes* (McGraw-Hill series on electrical engineering), McGraw-Hill.

Pareto, V. (1897) *Cours d'Economie Politique*, F. Rouge, Lausanne.

Paris, S. J. and Houthakker, H. S. (1955) *The Analysis of Family Budgets*, Cambridge, Cambridge University Press.

Pensions Commission (2005) *A New Pension Settlement for the Twenty-first Century*, second report of the Pensions Commission, London: TSO.

Percival, R. (2007) APPSIM – Software Selection and Data Structures, NATSEM working paper No. 3, National Centre for Social and Economic Modelling (NATSEM), University of Canberra.

Petrin, A. (2002) 'Quantifying the Benefits of New Products: The Case of the Minivan', *Journal of Political Economy*, **110**, 705-29.

Potter, M. (2010) Unpublished Ph.D. thesis.

Pudney, S. and Sutherland, H. (1994) 'How reliable are microsimulation results? : An analysis of the role of sampling error in a U.K. Tax-benefit model', *Journal of Public*

Economics, 53(3) 327-365, March 1994.

Railsback, S. F., Lytinen, S. L. and Jackson, S. K. (2007) 'Agent-based Simulation Platforms: Review and Development Recommendations', *Simulation*, vol. 82, no. 9, 609-623, September 2006.

Railsback, S. and Grimm, V. (2011) *Agent-Based and Individual-Based Modelling: A Practical Introduction*, Princeton University Press.

Rao, A. S. and Georgeff, M. P. (1991) Modelling rational agents within a BDI architecture, in Fikes, R. and Sandwell, E. (eds.) *Proceedings of Knowledge Representation and Reasoning (KR & R-91)*, 473-484, San Mateo, CA, Morgan Kaufmann Publishers.

Ravulaparthy, S, and Goulias, K. G. (2011) Forecasting with Dynamic Microsimulation: Design, Implementation and Demonstration, University of California Transportation Centre, UCTC-FR-2011-07.

Reynolds, C. (2008) Boids: Background and Update.
http://www.red3d.com/cwr/boids/

Rosenbaum, P. R. and Rubin, D. B. (1983) Biometrika, (1983) 70 (1): 41-55.

Rubin, D. B. (1987) *Multiple Imputation for Nonresponse in Surveys,* New York: John Wiley & Sons, Inc.

Ruiz, Y. & Walling, A. (2005) Home-based working using communication technologies, National Statistics Feature, Office for National Statistics, Labour Market Trends, October 2005.
http://www.statistics.gov.uk/articles/labour_market_trends/teleworking_Oct05.pdf

Rutter, C. M., Zaslavsky, A. M. and Feur, E. J. (2011) 'Dynamic Microsimulation Models for Health Outcomes: A Review', *Med Decis Making*, 2011, Jan-Feb, 31(1).

Samuelson, P. (1938) 'A Note on the Pure Theory of Consumer's Behaviour', *Economica*, **5**:61-71.

Samuelson, P. (1948) 'Consumption theory in terms of revealed preference', *Economica*, **15**:243-253.

SAS (2011) 'Estimating an Almost Ideal Demand System Model', SAS Institute Inc.
http://support.sas.com/rnd/app/examples/ets/aids/index.htm

Sauerbier, T. (2002) 'UMDBS – A new tool for dynamic microsimulation', *Journal of Artificial Societies and Social Simulation*, 5(2).
http://jasss.soc.surrey.ac.uk/5/2/5.html

Schelling, T. (1969) Models of segregation, *American Economic Review, Papers and Proceedings* 59 488-493.

Schmidt, B. (2000) *The modelling of human behaviour*, Erlangen, Germany: SCS Publications.

Schofield, D. (1995) Designing a User Interface for a Microsimulation Model, *STINMOD Technical Paper No. 7*, March 1995.

Scott, A. (2003) 'Implementation of demographic transitions in the SAGE Model', London, *ESRC-SAGE Technical Note no. 5*, London School of Economics.

Scott, J. and Carrington, P. J. (2011) *The SAGE Handbook of Social Network Analysis*, SAGE.

Sefton, T. (2007) Using the British Household Panel Survey to explore changes in housing tenure in England, Centre for Analysis of Social Exclusion, London School of Economics.

Smith, S. K., Tayman, J. and Swanson, D. A. (2002) 'Overview of the Cohort-Component Method', in Smith, S., K., Tayman, J., and Swanson, D., A. (2002) *State and Local Population Projections,* The Springer Series on Demographic Methods and

Population Analysis, 2002, 43-48.

Society of Actuaries (2007) Chapter 5 Corsim, Schaumburg, IL, USA, Table 5-1. Retrieved from http://www.soa.org/files/pdf/Chapter_5.pdf

Sommerville, I. (2006) *Software Engineering*, 8th Edition, Chapter 22.

Spielauer, M. (2007) 'Dynamic Microsimulation of Health Care Demand, Health Care Finance and the Economic Impact of Health Behaviours: Survey and Review', *International Journal of Microsimulation* (2007) 1(1) 35-53.

Statistical Consulting Group (2007) 'Introduction to SAS', UCLA: Academic Technology Services, Statistical Consulting Group, from http://www.ats.ucla.edu/stat/sas/notes2/

Statistics Canada (2008) *Modgen Developer's Guide* v9.0.10

Statistics Canada (2011) Modgen Models. http://www.statcan.gc.ca/microsimulation/modgen/new-nouveau/mods/mods-eng.htm

Stone, R. et al. (1954) *The Measurement of Consumer's Expenditure and Behaviour in the United Kingdom, 1920-1938*, Vol. 1, Cambridge: Cambridge University Press.

Stone, R. (1954) 'Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand', *The Economic Journal*, Vol. 64, No. 255 (Sep. 1954, pp511-527.

Studenmund, A. H. (2001) *Using Econometrics: A Practical Guide*, Fourth Edition, Addison Wesley Longman, Inc.

Sutherland, H. and Figari, F. (2013) 'EUROMOD: the European Union tax-benefit microsimulation model', *International Journal of Microsimulation*, 6(1) 4-26. http://microsimulation.org/IJM/V6_1/2_IJM_6_1_Sutherland_Figari.pdf

Symons, E. (1991) 'Household Microeconomic Simulation Models: A Short Survey', *The Economic Journal,* Vol. 101, No. 406 (May, 1991), pp 659-672.

Symons, E. and Walker, I. (1988) *SPIT version 4.0: simulation program for indirect taxation*, London, Institute for Fiscal Studies.

Symons, E. and Warren, N. (1996) 'Modelling consumer behavioural response to commodity tax reforms in micro-simulation models' in A. Harding (ed.) *Micro-simulation in Public Policy*, Elsevier Science Publishers, Amsterdam.

Takeuchi, K. (2009) Outlook for Retail Trade Sales – A Consumption Expenditure Approach Based on Demographic Trends, Financial Research Group, NLI Research.

http://www.nli-research.co.jp/english/economics/2009/eco090804.pdf

Tanton, R. and Edwards, K. L. (eds.) (2013) *Spatial Microsimulation: A reference Guide for Users*, Springer.

Taylor, M. (ed.) with Brice, J., Buck, N. and Prentice-Lane, E. (2009) *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*, Colchester: University of Essex.

Tedeschi, S. (2011) T-DYMM: Background and Challenges, Intermediate Conference.

Tesfatsion, L. (2002) Agent-Based Computational Economics: Growing Economies from the Bottom Up, *ISU Economics Working Paper No. 1*, 15 March 2002, Department of Economics, Iowa State University.

Tesfatsion, L. and Judd, K. (eds.) (2006) *Handbook of Computational Economics Volume 2: Agent-Based Computational Economics*, North Holland: Elsevier.

Theil, H. (1965) 'The information approach to demand analysis', *Econometrica* 33, 67-87.

Thomas, R. L. (1987) *Applied Demand Analysis*, Longman, London.

Tobias, R. and Hoffman, C. (2004) 'Evaluation of free Java-libraries for social-scientific agent based simulation', *Journal of Artificial Societies and Social Simulation*, 2004;7. <http://jasss.soc.surrey.ac.uk/7/1/6.html>

Trigg, A. B., Bertie, A. J. and Himmelweit, S. F. (2008) Modelling Bourdieu: An Extension of the Axelrod Cultural Diffusion Model, *CRESC Working Paper Series*, Working Paper 49, Open University, February 2008.

Turing, A. (1937) 'On Computable Numbers, With an Application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society*, Series 2, Volume 42 (1936-37) pp230-265.

Uhrig, S. C. (2008) The Nature and Causes of Attrition in the British Household Panel Survey, Institute for Social and Economic Research, University of Essex.

U.N. (2011) Classification of Individual Consumption by Purpose, United Nations Statistical Division. http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5

University of Essex, Institute for Social and Economic Research, British Household Panel Survey: Waves 1-17, 1991 – 2008 [computer file]. 5th Edition. Colchester, Essex: UK Data Archive [distributor], March 2009. SN 5151.

U.S. Census Bureau (2006) Survey of Income and Program Participation, retrieved from http://www.bls.census.gov/sipp/intro.html

Vargas, E., Schreiner, D., Tembo, G. and Marcouiller, D. (2011) 'Computable General Equilibrium Modelling for Regional Analysis', *The Web Book of Regional Science*, Regional Research Institute, West Virginia University. http://www.rri.wvu.edu/WebBook/Schreiner/chapter3.htm

Varian, H. (1982) 'The nonparametric approach to demand analysis', *Econometrica*, **50**, 945-972.

Varian, H. (2006) 'Revealed Preferance' in M. Szenberg, L. Ramaratton and A. Gottesman (eds.) Samuelsonian Economics and The 21st Century, Oxford University Press.

Venables, M. and Bilge, U. (1998) 'Complex adaptive modelling at J. Sainsbury: the SimStore supermarket supply chain experiment', The LSE Strategy & Complexity Seminar, 30 March 1998.

Vencatasawmy, C. P., E. Holm, T. Rephann, J. Esko, N Swan, M. Ohman, M. Astrom, E. Alfredsson, K. Holme and J. Siikavaara (1999) 'Building a spatial microsimulation model', SMC Internal Discussion Paper. Kiruna: Spatial Modelling Centre.

Voas, D. and Williamson, P. (2000) 'An Evaluation of the Combinatorial Optimisation Approach to the Creation of Synthetic Microdata', *International Journal of Population Geography*, 6: 349-366.

Wanless, D. (2002) *Securing Our Future Health: Taking a Long-Term View,* Final Report, HM Treasury.

Weintraub, E. R. (2002) 'Neoclassical Economics', *The Concise Encyclopedia of Economics*, Library of Economics and Liberty. http://www.econlib.org/library/Enc1/NeoclassicalEconomics.html

Westerhoff, F. and Franke, R. (2012) Agent-based models for economic policy design: two illustrative examples, *Working Paper No. 88*, BERG Working Paper Series, Bamberg Economic Research Group, Bamberg University, November 2012.

Whistler, D. (2008) 'Ordinary Least Squares Regression', An Introductory Guide to SHAZAM. http://shazam.econ.ubc.ca/intro/ols.htm

Wilensky, U. (1999) NetLogo, http://ccl.northwestern.edu/netlogo/, Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Williamson, P. (2009) 'Microsimulation in the UK: the current state of play',

presentation at the *ESRC Microsimulation Seminar Series,* Seminar 4:Bridging the Gaps; Setting the Agenda, University of Sussex, 11 September 2009. http://pcwww.liv.ac.uk/~william/ESRC-BSPS%20MSM%20Seminar %20Series_files/Williamson2%20edited%20transcript.pdf

Williamson, P., Birkin, M. and Rees, P. H. (1998) 'The Estimation of Population Microdata by Using Data from Small Area Statistics and Samples of Anonymised Records', *Environment and Planning* A, 30 (5): 785-816.

Williamson, P., Zaidi, A. and Harding, A. (2009) 'New Frontiers in Microsimulation Modelling', *Introduction*, in Zaidi, A., Harding, A., and Williamson, P., (2009) *New Frontiers in Microsimulation Modelling*, Ashgate Publishing Limited, UK.

Wilkinson, K. (2009) Adapting EUROMOD for use in a Developing Country – the Case of South Africa and SAMOD, EUROMOD Working Paper No. EM5/09, ISER UK. http://www.iser.essex.ac.uk/publications/working-papers/euromod/em5-09.pdf

Winder, N. (2000) Modelling within a thermodynamic framework: a footnote to Sanders (1999) *Cybergo: European Journal of Geography.*

Wing, I. S. (2004) 'Computable General Equilibrium Models and Their Use in Economy-Wide Policy Analysis', MIT Joint Program on the Science and Policy of Global Change, *Technical Note No. 6,* September, 2004.

Wolf, D. A. (2001) 'The Role of Microsimulation in Longitudinal Data Analysis', *Special Issue on Longitudinal Methodology, Canadian Studies in Population*, Vol. 28(2), pp313-339.

Wolfram, S. (2002) *A New Kind of Science*, Wolfram Media, Champaign, IL.

Wolfson, M. (2009) 'Preface – Orcutt's Vision 50 Years On', in Zaidi, A., Harding, A., and Williamson, P., (eds.) (2009) *New Frontiers in Microsimulation Modelling,* European Centre Vienna, Ashgate Publishing Ltd.

Wu, B. M., Birkin, M. H. and Rees, P. H. (2008) 'A spatial microsimulation model with student agents', *Journal of Computers, Environment and Urban Systems*, 32, 440-453.

Wu, B. M., Birkin, M. H. and Rees, P. H. (2010) 'A dynamic microsimulation with agent elements for spatial demographic forecasting', *Journal of Social Science Computing Review*, 29 (1), 145-160.

Zaidi, A. (2007) 'Model 3: The SAGE Model: A Dynamic Microsimulation Population Model for Britain', in Ann Harding (ed.) *Modelling Our Future: Population Ageing, Health and Aged Care (International Symposia in Economic Theory and Econometrics, Volume 16)*, Emerald Group Publishing Limited, pp.443-

447.

Zaidi, A. and Rake, K. (2001) 'Dynamic Microsimulation Models: A Review and Some Lessons for SAGE', SAGE Discussion Paper no. 2, SAGEDP/02.

Zaidi, A., Harding, A., and Williamson, P., (2009) *New Frontiers in Microsimulation Modelling*, Ashgate Publishing Limited, UK.

Zylberstajn, E., Zylberstajn, H., and Afonso, L. (2011) BRALAMMO: A dynamic labor market microsimulation model for Brazil.